PACIFIC LINGUISTICS

Series C - No. 93

FOCAL I: PAPERS FROM THE FOURTH INTERNATIONAL CONFERENCE ON AUSTRONESIAN LINGUISTICS

*

Paul Geraghty
Lois Carrington
S.A. Wurm
eds

Department of Linguistics
Research School of Pacific Studies
THE AUSTRALIAN NATIONAL UNIVERSITY

Geraghty, P., Carrington, L. and Wurm, SA. editors. *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics.*
C-93, x + 413 pages. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.cover*

PACIFIC LINGUISTICS is issued through the Linguistic Circle of Canberra and consists of four series:

SERIES A - Occasional Papers
SERIES B - Monographs
SERIES C - Books
SERIES D - Special Publications

All correspondence concerning PACIFIC LINGUISTICS, including orders and subscriptions, should be addressed to:

The Secretary
PACIFIC LINGUISTICS
Department of Linguistics
Research School of Pacific Studies
The Australian National University
Canberra, A.C.T. 2601
Australia.

First Published 1986
Typeset by Sue Tys
Printed by A.N.U. Printing Service Bound by Adriatic Bookbinders Pty Ltd
Maps drawn by Theo Baumann, Cartography, Department of Linguistics, Research School of Pacific Studies, Australian National University.

The editors are indebted to the Australian National University for assistance in the production of this series.
This publication was made possible by an initial grant from the Hunter Douglas Fund.

National Library of Australia Card Number and ISBN for complete set of two volumes: 0 85883 343 3

National Library of Australia Card Number and ISBN for this volume: 0 85883 344 1

TABLE OF CONTENTS

EDITORS' FOREWORD

As the series of International Conferences on Austronesian Linguistics continues to bring together Austronesianists whose interests range both linguistically and geographically across the entire field of this huge language family, so do the ensuing publications try to reflect the best thinking of each conference, the most innovative ideas that were presented, and not least the thrust and parry of discussion, which in numbers of instances has resulted in improvement to the papers themselves, between delivery and getting into print.

The fourth conference, held in Suva in August 1984, provided a wealth of material - over 60 papers - of which more than two-thirds have finally been included in these two volumes. In view of its rare aptness, the editors have chosen to retain the conference logo, and so the volumes are known as FOCAL I and FOCAL II.

Papers have been assigned to the two volumes along these lines: FOCAL I: syntax, pragmatics, and sociolinguistics; FOCAL II: historical and comparative linguistics, and language contact.

Production of the FOCAL volumes has not been without its difficulties, partly in that the chief editor was in Suva, the executive editor in Canberra, the typesetter in North Queensland for much of the time and then in Tasmania, whilst the advisory editor seemed to be more often in Peking, Palermo or Perú than in Canberra! But, thanks to the magnificent co-operation of the contributors, the volumes have been produced in not too unreasonable a time.

Some special thanks are due; first among all, the editors thank their typesetter par excellence, the imperturbable, perceptive Sue Tys, for her work; Theo Baumann's maps enhance the volumes; Tevita Nawadra has given us much encouragement, as have various linguists scattered across the Pacific and South-East Asia; many a modest photographer lent his prints for us to choose from; and we have to thank numerous photocopyists, printers and people in Canberra and in Suva, for their kindly tolerance towards FOCAL I and II.

PAUL GERAGHTY
Fijian Dictionary Project
Office of the Prime Minister
Suva

LOIS CARRINGTON
Department of Linguistics RSPacS
Australian National University
Canberra

S.A. WURM
Department of Linguistics RSPacS
Australian National University
Canberra

FOCAL REMINISCENCES

Memories linger longer when visits are shorter. Such was FOCAL (Fourth International Conference on Austronesian Linguistics). The week of August 13th to 18th, 1984 has long since gone but is not forgotten. So much happened in so short a time in so small a place as Suva. Never before had Viti hosted a linguistic conference, let alone one which attracted a truly international participation, despite distances involved. There was a spirit of togetherness, as participants came to meet, live, learn, and exchange views with each other and with fellow Austronesianists in Viti.

On behalf of the FOCAL Committee and all involved in Viti, I should like to say *vinaka saka vakalevu na yaco mai*. Many new friends were made, and old acquaintances renewed. As FOCAL was a conference on words, multilingual Suva was a most appropriate venue, and the participants were welcomed by none other than the *itaukei ni vanua kei na vosa* (owners of the land and its language) - and from then on, you were no more foreigners than they. We wish to thank the Tui Suva: *sā maleka vakalevu na Tūraga na Tui Suva kei na nomunū yavusa o Nadonumai kei Navakavu, e na nomunū vīmāroroi kei na vīqaravi*.

The conference papers ranged far and wide throughout the entire Austronesian area and aroused much interest and discussion, but not without refutation from some quarters. Such was the tone set by the two 'wise men', the guest speakers, *o rau na matua*, Dr G.B. Milner, and Dr G.W. Grace. Our wish is that your wisdom remain with us, *mo drau bota tu ga ka sega ni mago mai*.

We also wish to say *vinaka* to Professor S.A. Wurm and his steering committee for their advice which enabled the FOCAL Committee to function - *me dei tu ga na mata ni veisokotaki*; and we say *dhan badh* to the management and staff of the President Hotel whose good-will and generosity helped enormously in the running of the conference.

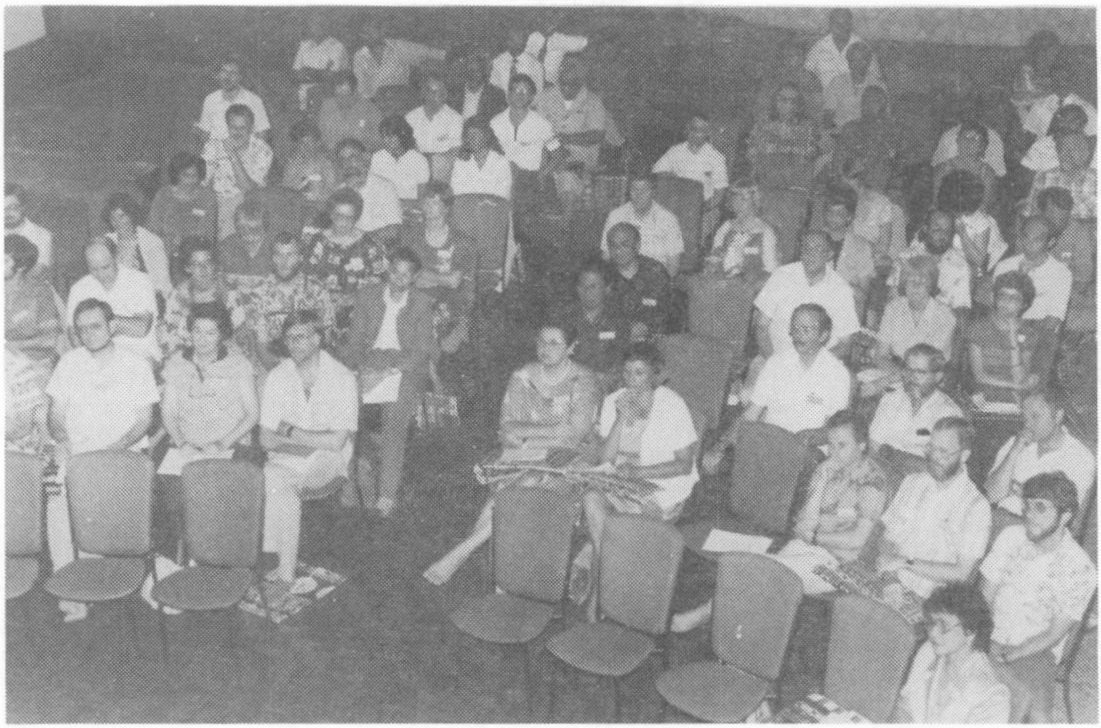
Time to say *moce* came and other duties across the waters had to be attended to, but Mount Rāmā will always be here like a beckoning hand. Parting was not the end, only the beginning of a new meeting. But for now, Viti says *moce, mo ni kalougata!* With you across the waters where the *waqa* of our Austronesian forespeakers once sailed, goes our song of farewell, always hopeful for a fresh reunion.

Isa lei, na noqu rarawa,	(Alas, how my heart is saddened,
Ni o sa na vodo e na mataka,	As you are leaving in the morning,
Bau nanuma na noḍatou lasa	Please remember our time together
Mai Suva, nanuma tiko ga!	In Suva, remember for ever!)

TEVITA R. NAWADRA

vinaka saka vakalevu na yaco mai - thank you very much for coming; *sā maleka vakalevu na Tūraga na Tui Suva kei na nomunū yavusa o Nadonumai kei Navakavu, e na nomunū vīmāroroi kei na vīqaravi* - thank you very much Tui Suva and your people of Nadonumai and Navakavu for your hospitality and kindness; *o rau na matua* - the two old men (or elders); *mo drau bota tu ga ka sega ni mago mai* - may you remain for ever strong and never wither; *me dei tu ga na mata ni veisokotaki* - may the eye that sets the course never falter; *dhan badh* - thank you (Hindi); *moce, mo ni kalougata* - farewell and good luck; *waqa* - canoe, vessel.

FOCAL: the First Day



Photograph: Ministry of Information, Suva

FOCAL: the Last Day

Photograph: David Zorc

FOCAL I: THE CONTRIBUTORS

DAVID G. ARMS, who comes from Lower Hutt, New Zealand, is a Catholic priest and a member of the Missionary Society of St Columban. After a few years working in Fiji, he completed a doctorate in linguistics at the University of Michigan. He is currently involved in Church translation work in Fiji; his main linguistic interests are in synchronic grammar, especially in Fiji and southern Vanuatu.

NIKO BESNIER received training in linguistics at the University of Southern California, where he is currently completing a doctorate, and Stanford University. He has conducted field research in various island groups of Western Polynesia and Melanesia, including, since 1980, Tuvalu. His interests are focused on the ethnography of speaking, sociocultural foundations of linguistic styles and discourse structure, and syntactic typology. He is also compiling a dictionary of the Tuvaluan dialects for the Tuvalu Government.

LOIS CARRINGTON has had wide experience in publishing, and in teaching English as a foreign language, in Europe, Australia and Papua New Guinea. In recent years she has been much involved in editorial work for *Pacific Linguistics*, and other research duties in the Department of Linguistics, Australian National University. She is a graduate of the University of Melbourne, in languages and history, and has undertaken postgraduate studies in education and Indonesian.

SUSANNA A. CUMMING comes from Boston, Massachusetts. She did her undergraduate work in linguistics at the University of California at Santa Cruz, and is currently working on her doctorate at the University of California, Los Angeles. She did her BA and MA work on Mandarin Chinese, addressing issues on syntactic change and clause combining, respectively; in her dissertation she is seeking to combine her interests in word-order change and functional syntax/discourse analysis in the context of the shift from predicate-initial to argument-initial in Austronesian languages.

OTTO CHR. DAHL was born in 1903 at Namsos, Norway; he served as an ordained missionary in Madagascar 1929-1957, thereafter in the administration of the Norwegian Missionary Society till 1966. In 1952 he completed a doctorate in linguistics at the University of Oslo. In 1967-1974 he had a government scholarship for linguistic studies and was then able to use his time for his main interests, Malagasy and Proto-Austronesian linguistics. He has been retired since 1975, but is still working.

VIDEA P. DE GUZMAN is Associate Professor in the Department of Linguistics, at the University of Calgary, Alberta, Canada. She received her PhD from the University of Hawaii. Her research interests are in the morphology and syntax of Philippine and other related languages.

CAROL P. GEORGOPOULOS grew up and went to college in Massachusetts, USA. She received her MA in linguistics from the University of Utah in 1980 and her PhD in linguistics from the University of California at San Diego in 1985. Written under the guidance of Professor Sandra Chung, her dissertation explores the syntax of unbounded dependencies in Palauan. She is currently an Assistant in Linguistics at the University of Geneva, Switzerland.

PAUL GERAGHTY was born in England of Irish parents, and educated at Rugby School and Cambridge University. His doctoral dissertation, from the University of Hawaii, was on the history of the Fijian languages. He has been working since 1978 as Consultant/Researcher at the Fijian Dictionary Project, in Suva.

VOLKER GRAEFE comes from Burg, Germany. After studying in Göttingen, Munich, Hamburg and Kiel, he completed his PhD in Physics at the University of Kiel. He was an Assistant Geophysicist in Oceanography at the University of Hawaii, and a group leader in the Krupp Atlas-Elektronik company in Bremen, Germany. He is currently Professor of Measurement Science in the Aerospace Department of the Federal Armed Forces University, Munich. His main interests are image processing, pattern recognition and artificial intelligence.

RAY HARLOW comes from Kent, UK, but has lived in New Zealand for many years. He completed a doctorate in Zurich with a dissertation on Greek dialects of the third century BC, and with a magna cum laude examination in Greek, Latin, and Indo-European, and taught Classics in New Zealand for some years. Since 1977, he has been Senior Lecturer in Linguistics at the University of Otago, Dunedin; his main research interests lie in the area of Maori and the Eastern Polynesian languages, their description and interrelationships.

ROBIN HOOPER comes from Auckland, New Zealand. She teaches English language at the University of Auckland, where she is a PhD candidate in Linguistics. For several years she has acted as adviser and editor to the Tokelau Dictionary Project. Her research concerns comparative Polynesian grammar and the language of Tokelau.

BAMBANG KASWANTI PURWO comes from Jakarta, Indonesia. He is currently Senior Lecturer in General Linguistics at the Catholic University of Atma Jaya, Jakarta, an editor for the series NUSA: Linguistic Studies in Indonesian and Languages in Indonesia, and secretary to the Linguistic Society of Indonesia (1985-1988). He was a postdoctoral research fellow at the Institute for Advanced Study, Princeton, New Jersey, in 1983-1984. His main interests are syntax, semantics, and (discourse) pragmatics.

PAUL R. KROEGER is a native of St Paul, Minnesota. He holds an MA in linguistics from the University of Texas at Arlington, and is a member of the Malaysia Branch of the Summer Institute of Linguistics. He is currently engaged in research on the Kimaragang Dusun language of Sabah, East Malaysia.

JULIETTE LEVIN is presently assistant professor of linguistics at the University of Texas at Austin, and was educated at Barnard College and the Massachusetts Institute of Technology. Her current research projects include further investigation of surface ergativity, typological research on distributive morphology, and studies in metrical phonology.

DIANE MASSAM is presently lecturer in linguistics at the University of British Columbia, and was educated at York University, the University of Toronto, and the Massachusetts Institute of Technology. Her current research projects include cross-linguistic analysis of exceptional case-marking structures, with specific attention to Austronesian languages, as well as syntactic analysis of possessor-raising constructions.

GEORGE B. MILNER retired in 1982 from the School of Oriental and African Studies at the University of London, where he spent most of his career, which was interspersed with several years of fieldwork in the Western Pacific, mostly in Fiji and Samoa. He is the author of a Fijian Grammar and of a Samoan Dictionary and he is at present working on a new Fijian-English Dictionary.

SUSAN MORDECHAY is completing her doctoral dissertation in linguistics at the University of California, Los Angeles. Her published work so far includes papers in semantics, syntax, and discourse analysis with an areal emphasis in Austronesian, or rather, Indonesian languages.

PAZ BUENAVENTURA NAYLOR is Assistant Professor of Linguistics at the University of Michigan, Ann Arbor, teaching courses in Linguistics and Tagalog. She is much interested in topic, focus, and related questions, and in Austronesian linguistics in general.

STANLEY STAROSTA is Professor of Linguistics at the University of Hawaii. He has done fieldwork on Austronesian languages in Taiwan and on Munda languages in southern India. His main theoretical interests are in syntax.

HEIN STEINHAUER has a doctorate in Slavistics from the University of Amsterdam. In the mid-seventies he did fieldwork on the non-Austronesian Blagar language, Alor archipelago, Indonesia. Since then he has been mentoring Indonesian linguists and has done fieldwork on several Indonesian languages within the framework of the Indonesian Linguistics Development Project. He is at present a lecturer in Austronesian linguistics in the Department of Language and Cultures of South-East Asia and Oceania, University of Leiden, Netherlands. His main interest is descriptive linguistics.

GERALDINE TRIFFITT was born in Tasmania and came to Canberra with a degree in geography and German. She has since worked as a geographer and as a librarian in several government instrumentalities, including twelve years at the National Library of Australia. After her family was adopted into a Fijian village in the Yasawa Islands, she completed a part-time degree in linguistics at the Australian National University with a view to studying the Western Fijian language.

SUE TYS is senior of the Linguistics Assistants in the Department of Linguistics, Research School of Pacific Studies, at the Australian National University. She has been responsible for much of the updating of production style for *Pacific Linguistics* in recent years, and has typeset many of the major *Pacific Linguistics* publications.

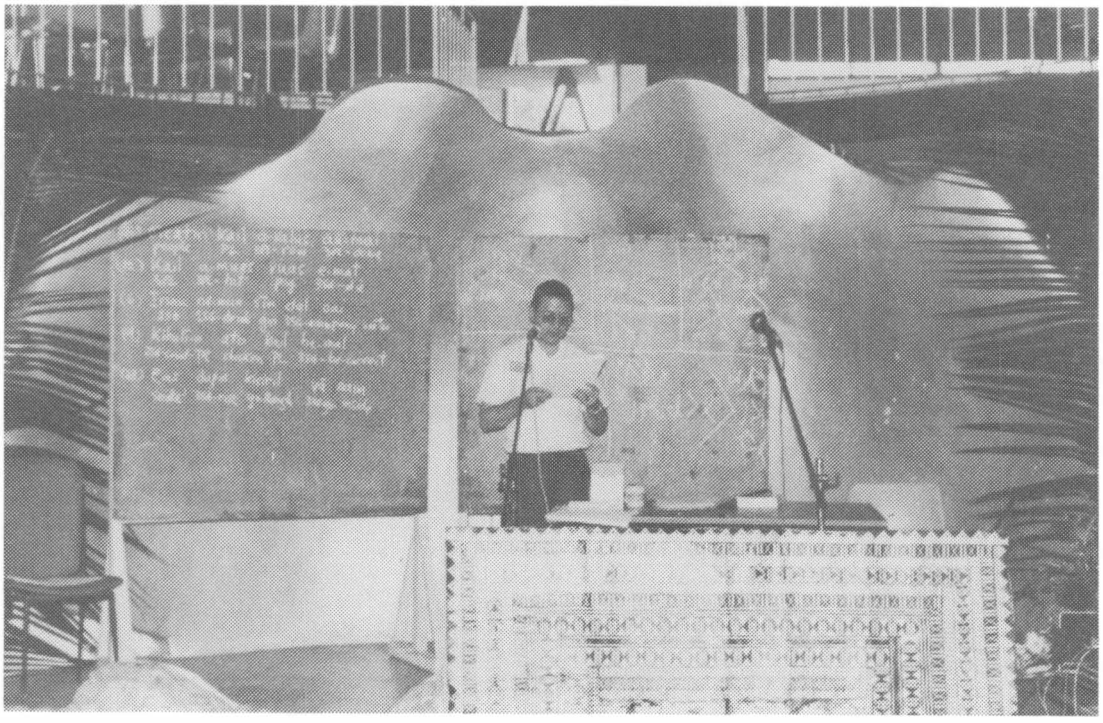
FAY WOUK is a graduate student in the Department of Linguistics at the University of California at Los Angeles. She is an Indonesianist and discourse analyst who has done discourse work on Toba Batak, Tagalog and Indonesian, and is currently working on a dissertation on the discourse distribution of verbal morphology in the Standard Indonesian of Jakarta.

STEPHEN A. WURM is Professor of Linguistics and Head of the Department of Linguistics, Research School of Pacific Studies, Australian National University, Canberra. His main research interests are concerned with the language situation and sociolinguistic questions of the Greater Pacific Area, including South-East Asia and China, with the south-western Pacific and its Papuan and Austronesian languages being his area of foremost interest, on which he has published widely.

AYAKO YASUDA-GRAEFE comes from Nagoya, Japan. After studying anthropology and linguistics in Japan and USA, including field studies in New Zealand, she completed her MA in linguistics at the University of Hawaii. She has lectured in Polynesian linguistics at the University of Copenhagen, Denmark, and at the University of Munich, Germany. Her main interests are Polynesian linguistics and German phonetics, especially intonation.

R. DAVID ZORC comes from North Chicago, Illinois; after work with the Peace Corps (1965-1969) in the Philippines, he undertook his doctorate at Cornell University (1969-1973) and worked as an Assistant in Research at Yale (1973-1975). Since 1976, he has been a Senior Lecturer in the School of Australian Linguistics of the Darwin Institute of Technology, in the Northern Territory of Australia. His main interests are lexicography, subgrouping, and historical reconstruction.

FOCAL: Session in progress



David Zorc

Geraghty, P., Carrington, L. and Wurm, SA. editors. *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics.*
C-93, x + 413 pages. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.cover*

A FOCAL APPROACH TO PROBLEMS OF VERBAL SYNTAX IN FIJIAN

G.B. Milner

One of the paradoxes of the present state of our studies is the contrast between the absolute confidence we have in the organic unity of AN phonological and lexical systems, and our curious reluctance to take full advantage of the fact that AN grammatical systems are also cognate and thus open to analogous procedures of linguistic analysis.

This may be due in part to the great development and considerable achievements of phonetic and phonemic studies during the present century, by contrast with the slower progress of our understanding of morphology and syntax. The renewed interest in linguistic typology and universals of grammar, in recent decades, may also account for a noticeable tendency to analyse AN languages with at least one eye (and sometimes both eyes) on better-known grammatical systems, particularly those of the Western European type like English.

It is also paradoxical and ironical, that the more we condemn traditional school grammar for applying inadequate and obsolete concepts such as the 'parts of speech', declensions and conjugations, which originally go back to the study of Latin and Greek, to non-Western languages, the less we notice our own tendency to write, not grammars of AN languages within their own terms of reference, but grammars of what they appear to be when translated into English and submitted to a framework of analysis established in the tradition of IE grammar.

This is not to say that in order to understand AN grammar, we need to abandon the cardinal points, the main parameters of our linguistic universe, such as verbals and nominals, predication, transitivity or tense and aspect. What is perhaps required of us, however, is that these concepts should be defined more flexibly and adjusted to the needs of Austronesian languages (instead of the reverse as happens so frequently). That is to say we should have the open-mindedness of the discoverer, as well as the courage of the explorer, leaving linguistic universals and typology on a slow burner, at least until we are satisfied that we really understand the fundamental principles of AN grammar.

Instead of starting from scratch in a neck of the AN woods, as I did many years ago with my little axe, without paying much attention to what greater men had achieved elsewhere before me, it seems to me that we should have more respect for our predecessors. There is after all one language, in the study of which a distinguished tradition of painstaking and thorough grammatical analysis has long been established. One thinks especially of the work of men like Bloomfield and Lopez, as well as others, in Tagalog. Because they achieved so much more than has so far been possible in other AN studies, what the scholars of Tagalog have to say seems to me to be of far greater relevance for the progress of our studies, than the linguistic fashions of the moment.

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 1-20. *Pacific Linguistics*, C-93, 1986.

The co-authors of a comprehensive reference work for instance, show that Tagalog, in addition to verbal predication, which they describe as "narrational sentences", has two types of nominal predication, which they call "equational sentences" (Schachter and Otanes 1972:61) and that

> It may be argued that the distinction ... between equational and narrational sentences ... is a somewhat arbitrary one, and that all Tagalog basic sentences ... are essentially equational in nature, involving a balance of two elements - the predicate and the topic - against one another (p.62)

In an instructive article published a few years later, Schachter (1976), states that it is "centrally concerned with the question of whether or not there are identifiable subjects in the sentences of Philippine languages ..." (p.493). His conclusion (p.513), is that "there is in fact no single syntactic category in Philippine languages that corresponds to the category identified as the subject in other languages". Moreover "if the conclusion ... is in fact correct, then obviously it cannot be the case that *subject* represents a linguistic universal" (p.515). More recently, in a renewed study of the same problem (1977a) he states that

> ... since the (Philippine) languages CAN be analysed quite satisfactorily as NOT having subjects, I conclude that the assumption of the universality of subjects is, in the case of Philippine languages, something of a Procrustean bed, and see nothing to be gained by forcing the languages into this bed.[1]

Just as in the fields of linguistic typology, and universals of grammar, a great deal of theoretical work has turned around the problems raised by the correct identification of subject, actor (or agent), object, goal and topic, and thus of active, passive, or ergative constructions, so in recent studies of Central Oceanic languages, preoccupation with case and voice (*diathèse*, to use a term employed by Tchekhoff (1978:37-57)), has been especially noticeable.[2]

Perhaps one should resist the temptation, at my age, to pose as a prophet and admit instead to the errors of youth. Let me therefore explain that there was a time when, having dismissed verbs and nouns, adjectives and adverbs, as being suitable terms for an adequate description of Fijian, I could yet write about the 'passive' (in spite of the fact that it had no 'agent'), as well as about subjects, objects and transitivity. Yet I was not (and still am not) alone. Biggs for example, in a remarkably shrewd and perceptive article (1974:404-405), where he breaks much new ground, gives his reasons for not abandoning the categories of IE grammar in PN languages. Thus he retains 'actor-subjects' as opposed to 'goal-subjects'. Schütz and Nawadra (1972:97-107), after throwing much revealing light on this vexed question, abandon 'passives' in Fijian but they leave us with 'participles'.

The difficult situation in which one can be trapped by an unguarded use of terms extrapolated from IE grammar, was well described by Clark in the appendix to his review of transitivity and case in Eastern Oceanic (1973:594-598). He accepts an active/passive distinction for some languages, but he leaves us in doubt about Maori (p.598). Surprisingly, he also questions the validity of distinguishing between 'passive' and 'ergative' (p.597). Yet if he does not entirely remove our uncertainty, he has earned our thanks by restating the difference between the two principal verbal constructions in PN (p.569; pp.574-575).[3]

In a wide-ranging survey of Proto-Oceanic grammar, Pawley (1973:116-119) considers subjects, objects and the order of constituents, as well as transitive suffixes and verb classification (pp.120-147). Much of his material is drawn from Fijian (especially pp.137-140) and we are in his debt for recognising the structural similarity between 'passive' and 'transitive' (p.137), and the possibility that the so-called 'passive' may be a subclass of the 'stative' (p.139).[4] Yet he leaves other problems of Fijian verbal syntax unresolved.

In 1976 at the Second Eastern Conference at Ann Arbor (Milner 1979), I suggested that topic and focus, which had been associated in the past mainly with Tagalog and Philippine linguistics, might offer a productive line of investigation for Oceanic languages. In the same volume of papers, Ferrell and Stanley (1979: 19-31) make a powerful plea for the relevance of topic and focus to the languages of Taiwan. In the same publication, Wolff (1979:155-160) shows striking parallels between the behaviour of noun and verb phrases in Samareño (Philippines), Arosi (Solomons) and Fijian, and he gives evidence for considering that Oceanic languages have focus rather than voice. Added weight was given to the view that focus might be a pan-Austronesian category by Dahl, at the SICAL conference in Canberra, when he endorsed the validity of the same approach in the case of Malagasy (1978:383-393). Closer to my own interests and on the same occasion, Naylor (1978:395-442), in a broadly-based survey of syntactic problems of several languages (selected from Micronesia, Polynesia and Melanesia), also made a strong plea for the need to make a new investigation of Oceanic grammar based on the solid, yet geographically limited, progress made in Philippine and in Taiwan linguistics.

'Focus' has been used by different linguists to refer to different concepts. As Naylor has stated (1975:16-17), some writers view focus as the highlighting of a noun, others as a type of emphasis. It is not, however, profitable to follow them, since in many AN languages, including Oceanic, emphasis in the clause has been correlated with initial or prepredicate position. On this point (1975: 16), she quotes Pike:

> Focus is not emphasis. Focus reports the observer's attention to one of several relations - without the essential emotional overtones - between a predicate and some other part of a clause; the focus-complement substantive topic is viewed only in reference to that relationship, not as in focus of itself. In emphasis, on the contrary, some one substantive is singled out for a direct isolated overlay of emotional connotation .... This formal independence of emphasis allows it to function as a variable which is formally separate from the focus complement.
>
> (Pike 1963:219)[5]

In a section of the article based on her Canberra paper, Naylor (1978:395-442) considers in detail the theoretical relevance of focus to Oceanic languages. Let me quote three short passages from it which need to be remembered:

> Although the surface forms that manifest focus vary from one AN language to another, the function is shared by all. It is a function that is central to the organisation of discourse in the general and specific senses ... (p.396)

> ... focus is like a prism; it has several facets. Not all AN languages overtly encode all of its facets in the syntactic structure. Even when the same facet is encoded, the manner of encoding varies. Furthermore one facet may be salient in one language but not in another. (pp.398-399)

> Topic-and-comment has ... been distinguished from subject-and-predicate. The former includes notions that are proper to sentence and discourse structure while the latter consists of notions that are proper to clause structure; i.e. topic-and-comment deals with extra-sentential relations while subject-and-predicate deals with sentence-internal relations ... (p.402)

In their contribution to the collected papers from the Second Eastern Conference, Ferrell and Stanley (1979:19-31), using material from Taiwan, present an argument for regarding focus in AN as an indispensable category. They illustrate it from Paiwan and their view is particularly interesting because they partly support those of Naylor and also introduce others with a different ground. Thus:

> ... focus in AN is not equivalent to case ... focus is a sort of topicalization whose selection has to do with discourse continuity .... *Strictly speaking, focus is not topicalization of one of the overt NPs of the sentence, but rather of one of the restricted number of underlying semantic-role categories which NPs may fulfill with reference to specific verbs* (their italics). This semantic-role focus is indicated by the obligatory focus inflection on the verb; the occurrence of an overt NP identifying or explicating the in-focus element is optional. (p.19)

Particularly striking, moreover, are their remarks on *focus inflections and nominals*, which immediately bring Fijian to mind:

> The semantic content of the verb focus inflections, as indicating agent, object, referent, or instrument, is seen in their use with nouns. (p.23) e.g.:
>
> kan-en (1) *food*, (2) *eat* (OF) (< kan *eat*, -en [object])
> si-kan (1) *eating utensil*, (2) *eat* (IF)
> si-kasuy *something used as trousers* (< kasuy *trousers*, si-[instrument])

One is at once reminded of similar possibilities in Fijian, as in:

vakabaut- *believe*
na vakabauta *faith, belief*

This feature of Paiwan is one of the essential criteria used by Ferrell and Stanley (p.29) to distinguish focus from case. That is to say, whereas

WITH FOCUS	WITH CASE
The inflected verb can stand as a noun itself (e.g. an IF-inflected verb may be used in a noun slot with the meaning *instrument used for*)	The inflected primary verb can never be used as a noun

This is a distinction which would also be valid for Fijian. To quote the same authors again (p.28): "It is failure to distinguish process from role that in the past has led investigators to confuse AN focus with IE voice, and more recently with IE case".

Another important point (made by Naylor) which is also familiar to students of Fijian and other Oceanic languages, is the following: "... the choice of focus makes a contrastive semantic difference: (thus) a non-definite vs. definite direct object (goal)" (1975:17).

The words "definite" and "non-definite" at once strike a chord for someone used to what has been called by Hazlewood (1872:32-33) and Churchward (1941:17-18; 71-72) "definite transitive" as opposed to "indefinite transitive" in Fijian, or also regarded as an "incorporated object" (Naylor 1978:419), as in the distinction, perhaps by now familiar to many, between for instance:

> gunu yaqona *kava-drinking* (or *grog-drinking* for old Fiji hands) and
> gunuva na yaqona *drinking the* (or: *some*) *kava* (Milner 1972:26; Wolff 1979:155; Naylor 1978:417-418; see also Hopper and Thompson 1979: 257-258 for a similar feature in Tongan)

Let me set out the arguments militating in favour of considering Fijian to have a focus-and-topic system of verbal syntax which, in its essentials, is analogous to that which has been described for Tagalog, but which, in its operation and methods of affixation, differs markedly from the latter. It is necessary, first, to remove the obstacles which have prevented us until now from seeing the wood, as distinct from the trees. That is to say, we must clear the decks by putting the numerous allomorphs of Fijian verbal suffixes into clearer perspective.

Our understanding of this old problem, which had already intrigued Hazlewood before 1850 and which is still with us, has been greatly advanced by several publications during the last 20 years, notably those of Krupa (1966); Arms (1973); Hockett (1976; 1977); Schütz (1981); Geraghty (1983); Clark (1977); and Lichtenberk (1978). It is Arms, however, in his Ph.D. dissertation (1975), who must get the main credit for spelling out in detail the operation of the phonotactic constraints on the occurrence of the thematic consonants in the verbal suffixes (1975:136-147).[6] Unfortunately, the importance of what he states on this subject has perhaps been masked by his suggestion that thematic consonants have semantic or 'phonesthetic' connotations. That is a view which, in certain instances, can be defended and which Hockett (1976; 1977; cf. Milner 1986:note 27; cf. also note 16 below) and Geraghty (1983:267-269) have supported, but one which it is very difficult to reconcile with the operation of phonotactic rules. The latter show remarkable rigour and consistency.

Many years ago Scott (1948:737-752) presented the first analysis of Fijian phonology by a professional phonetician. It includes a table (p.743), which illustrates the almost complete concordance between consonant graphemes and phonemes and evidence for four places of articulation: bilabial, dental, alveolar and velar.

In his Ph.D. dissertation, Arms (1975:136-147) shows that with hardly any exceptions, the place of articulation of any consonant in a Fijian verb base rules out the occurrence in the suffix of a thematic consonant with the same place of articulation. For example, v is ruled out if any one of the following occurs in the base: v, b, m or (subject to certain exceptions)[7] w.

Consonants which cannot occur together in given positions, are said to *dissociate*, or to be *dissociative* (cf. Krupa 1966; Arms 1975:130-147) and the phonotactic rules which can be established accordingly, can be called rules of dissociation.

In a recent article (Milner 1986), I have argued that the operation of those constraints has so far not been analysed with a view to discovering the principles which govern the occurrence of what can be called *heterorganic* or *replacement* consonants, i.e. those substituted for consonants which, but for a phonotactic rule, would normally occur and which I shall call *regular* thematic consonants. As a consequence, the effects of the replacements on the nature of the system of suffixation have been obscured.[8]

In actual fact, if we disregard replacement consonants, that is to say, if we treat them as allomorphs, which, by definition can be subsumed under morphemes, the essential features and principles of the system become much clearer than they have appeared to be until now.

The phonotactic rules can be stated fairly simply by placing consonants which dissociate from one another in the same columns of a table, as shown below. Consonants occurring in verbal suffixes are underlined:

b v m w	. c	d t	. l s	dr r n	q k g

Two important features should be stressed in connection with this table:

(1) In a *short* suffix (i.e. a monosyllabic suffix),[9] the thematic consonant may be Ø (zero). The suffix is then reduced to -a after a front vowel (and after the open vowel (except in taya *hit*). -ya occurs after a back vowel.

(2) -Ø- and -t- occur very frequently. Together, they account for 569 recorded suffixes in Arms' total list of 1680. They were also the two suffixes which he had the greatest difficulty in correlating with any special semantic connotations (Arms 1975:126; 110-112; 113-114).

As long ago as 1850 (in the first edition of his grammar), Hazlewood proposed the rule that: "... *verbs formed from nouns without prefixing* vaka-, ... shall take na for their termination ..." and "*verbs of motion* ... will take va for their termination" (1872:33). He also added that it was "also true that many other verbs besides those of motion take va, but for these perhaps there is no rule". (p.33).

It is possible, however, that phonotactic rules may determine the occurrence or non-occurrence of -v- as a thematic consonant in the following two ways:

(1) If a "verb of motion" (to use Hazlewood's phrase) has a bilabial consonant in its base, its thematic consonant will not be a *regular* -v-, but a *heterorganic consonant*, i.e. one which is not bilabial.

(2) Conversely, if a base has a -d- or a -t-, this will rule out the occurrence of a *regular* -t- in the suffix (assuming for the moment that some rule governs the occurrence or non-occurrence of -t- as well as of -v-). In that case the *heterorganic consonant* may be -v- or some other consonant.

From a fairly detailed study of verbal suffixes made on the evidence provided by three separate dictionaries,[10] it is possible to establish the following data:

1. Short verbal suffixes

(a) -Ø- and -t- occur as *regular* thematic consonants when the process or action is carried out near the actor, i.e. in relatively close physical or psychological space.[11] The term *static goal* is suggested to describe this type of suffix. (NB: There seems to be no significant difference of grammatical function between -Ø- and -t-. In general either one or the other is used with any particular base.[12] -Ø- frequently occurs when the second vowel of the base is -i-. Conversely, -t- frequently occurs when the second vowel of the base is open or back.)

(b) -v- occurs as a *regular* thematic consonant when the process or action is extended to, or is exercised over a distance, i.e. involves a displacement in physical or psychological space.[11] The term *kinetic goal* is suggested to describe this type of suffix.

(c) A number of minimal pairs can be established so as to validate the distinction proposed in (a) and (b) above.[13]

(d) HETERORGANIC CONSONANTS

(i) The following consonants occur most frequently as *replacements* when a *regular* thematic consonant would be homorganic with a consonant of the base:

Instead of -v-: -c-, -k-, -r- and -t-[14]
Instead of -t-: -c-, -k-, -r- and -v-[15]

(ii) The following consonants occur least frequently as *replacements* and may be subject to additional rules: -g-, -m- and -n-.

-g- is a rare heterorganic consonant.[16]
-m- is a rare heterorganic consonant.[17] It also occurs occasionally when the second vowel of the base is -u-.[18]
-n- is a rare heterorganic consonant.[19] It occurs mainly as a denominal verb suffix and in verbs which, except in derived forms, do not appear to occur without a suffix.[20]

(e) ADDITIONAL NOTES

(i) When a base with a bilabial consonant has a suffix with -t- as a thematic consonant, or

(ii) When a base with -t- or -d- has a suffix with -v- as a thematic consonant, or

(iii) When both a bilabial and -t- or -d- occur in the base, it is more difficult to determine whether a regular thematic consonant has been replaced, and if so, to decide which regular consonant is replaced by which other consonant. There is, however, some indication that -t- may be replaced by -c- and -v- by -r-.[21]

2. Long verbal suffixes

(a) The following consonants occur as thematic consonants of long suffixes: -c-, -k-, -l-, -m-, -n-, -r-, -t-, -v-, -y-.[22]

(b) Three of these have a special function:

(i) -y- often occurs in the suffix -yaki when a base is prefixed by vei-. It denotes that a process or action is exercised in different directions, at random, indifferently, etc.

(ii) -l- denotes that a process or action is exercised frequently and/or repeatedly.

(iii) -r- denotes that a process or action is exercised with force and/or violence and/or intensity.

(c) The occurrence of thematic consonants in *long suffixes* is subject to the same rules of dissociation which apply to *short suffixes*, except for -t-, which does not dissociate.[23]

(d) -t- occurs as a regular thematic consonant when the process or action is oriented towards:

(i) an instrument with which it is carried out or exercised,[24] or

(ii) an object affected by the process or action,[25] or

(iii) an object produced by the process or action.[26]

The term *instrumental* is suggested to describe this type of suffix.

(e) -v- occurs as a regular thematic consonant when the process or action is carried out or exercised:

(i) on account of, about, someone or something,[27] or

(ii) on behalf of, for (the benefit of), someone or something.[28]

The term *beneficiary* is suggested for this type of suffix.

(f) The distinction between instrumental suffix and beneficiary suffix can be validated by the existence of minimal pairs.[29]

(g) HETERORGANIC CONSONANTS

(i) -t- is a replacement consonant for -v- in accordance with the phonotactic rules already discussed for the short suffixes.[30]

(ii) -c-, -m- and -v- are replacement consonants for -t- as a result of what seems to be analogy with the thematic consonant which occurs in the short suffix of the same base.[31]

(iii) -k- and -r-: Only a few instances of -k- as a thematic consonant in long suffixes are attested. At least one instance of -r- without any suggestion of force and/or violence (see (b)(iii) above) is also known.[32] It is not possible at present to account for the occurrence of those two consonants in long suffixes.

(iv) -n- is also rare in long suffixes. In at least two cases it seems (like -na as a short suffix) to occur as a denominal verbal suffix.[33]

It is possible now to identify some of the main features of a system of focus-and-topic in the verbal syntax of Fijian. If the evidence before us is sound, as we have reason to believe, we are left with a relatively small number of morphemes which can be firmly linked with two types of *goal focus* (one *'static'* and the other *'kinetic'*), an *instrumental focus* and a *beneficiary focus*. In order to complete the pattern, however, it is necessary to look for an equivalent in Fijian of the *actor focus*, which in Tagalog and other Western Austronesian languages, can be identified formally.

Before doing so, let me first recall Naylor's remark that "focus is like a prism .... Not all AN languages overtly encode all of its facets in the syntactic structure" (1978:398-399). Ferrell and Stanley, for their part, state that "... the occurrence of an overt NP identifying or explicating the in-focus element is optional" (1979:19).

Biggs, on the other hand, has drawn our attention to the all-important distinction in Fijian between two classes of verbs, namely "those that choose an actor-subject and those which choose a goal-subject"[34] (1974:418). He returns to this point in the concluding paragraphs of the same article, arguing that one type of suffix "derives actor-subject selecting verbs from goal-subject selecting verbs" while another suffix "Conversely ... derives goal-subject selecting verbs from actor-subject selecting verbs" (p.425).

It would seem that this is a crucial distinction, which needs to be looked at again. Let me however put forward a different explanation for it.

I wish to propose that the principal distinction between these two classes of verbs, which Schütz (1981:201; 1986:112) calls *stative* (i.e. goal-selecting) and *active* (i.e. actor-selecting) is that the expression of focus in one class is the converse of its expression in the other, that is to say they are symmetrically reversible.

Instead of the terms used by Biggs and Schütz, let me first propose that we should speak on the one hand of

ACTOR-ORIENTED VERBS (AOV) (i.e. 'active' according to Schütz and 'actor-subject selecting' according to Biggs), e.g.

gunuv- *drink*, lakov- *go*, raic- *see*, kaniØ- *eat*

and on the other hand of

GOAL-ORIENTED VERBS (GOV) (i.e. 'stative' according to Schütz and 'goal-subject selecting' according to Biggs), e.g.

biut- *leave*, rogoc- *hear*, cakav- *do, make*, sogot- *close*

Stated briefly, the view put forward now is the following:

An AOV without a suffix is in *actor focus*

An AOV with a suffix is in *goal focus*

Conversely:

A GOV without a suffix is in *goal focus*

A GOV with a suffix is in *actor focus*

The last of these propositions will perhaps cause some surprise and perhaps controversy. A careful examination of the evidence, however, can only leave one convinced that no other interpretation of the data will do. Before giving supporting evidence, let me paraphrase the above statements.

If we speak of an ACTOR-ORIENTED VERB, we mean that the role of *actor* is already assumed (i.e. that it is implicit in, part and parcel of, a verb). It therefore does not require a formal exponent and the unsuffixed base (i.e. its unmarked form) is oriented towards the topic NP in the role of *actor*).

Therefore, in an AOV, a short suffix (i.e. a marked form of the base) is available and appropriate when the base is oriented towards the topic in the role of *goal*.

Conversely, if we speak of a GOAL-ORIENTED VERB, we mean that the role of *goal* is already assumed (i.e. it is implicit in, part and parcel of, a verb). It therefore does not require a formal exponent and the unsuffixed base (i.e. its unmarked form) is available and appropriate when the base is oriented towards the topic NP in the role of *goal*.

Therefore, in a GOV, a short suffix (i.e. a marked form of the base) is available and appropriate when the base is oriented towards the topic NP in **the** role of ACTOR.

This last statement will require some theoretical support as well as corroborating evidence.

In the grammatical tradition associated with the study of IE languages, and particularly of Latin and Greek, it has been customary to begin with small segmentary units such as phonemic and morphemic constituents, and only then to study phrases, clauses and sentences. As a result, complete utterances (especially VPs in syntagmatic association with two, three or even four NPs) have received more attention than their frequency of occurrence in spoken (as opposed to written, especially literary) language, would seem to justify. That is to say, we have tended to study model constructions (elicited from informants and divorced not only from their socio-physical environment and subject of discourse, but from their linguistic context) at the expense of living speech. The latter, of course, takes full advantage of situation, shared knowledge and contextual information.

If therefore we attempt to elicit complete sentences or utterances from our AN-speaking informants, as Biggs (1974:401-408) has advocated, we run the risk of overlooking the important role played by anaphora within discourse. Let me illustrate this point from three Fijian riddles and three proverbial expressions (Biggs 1948; Bulicokocoko c.1957; Anon. n.d.). These have been deliberately chosen as examples since, within a given socio-cultural and socio-physical environment, a riddle and a proverb may refer to any subject of discussion whatever, the only prerequisite condition being that speaker and hearer(s) should share the same knowledge.

FIRST RIDDLE: 'Who is the visitor who always comes twice, in the day-time and at night-time?' Answer: 'The tide' (Biggs 1948:43)

O cei na vūlagi // ka dau lako mai vakarua ...
ANAPHORIC ACTOR TOPIC AOV
Who (is) the visitor who always comes twice ...

That is to say lako, an ACTOR-ORIENTED VERB, *without* a suffix, is in ACTOR FOCUS, i.e. oriented towards a NP in the role of ACTOR. ka dau lako mai vakarua (*who always comes twice*) is a relative clause, with anaphoric relation to na vūlagi (*the visitor*). When the relative particle ka occurs, the third person singular pronoun in preverbal position is Ø (i.e. 'deleted'). The TOPIC NP here is thus absent. (It will be remembered that "the occurrence of an overt NP identifying or explicating the in-focus element is optional" (Ferrell and Stanley 1979:19).)

SECOND RIDDLE: 'There is a pond full of water. A white cloud forms (in the middle of it) which drinks up the pond'. Answer: 'A coconut: When it's about to germinate, the milk dries up because the pith absorbs it' (Biggs 1948:343, no.10).

... e dua na ō vulavula ... ka Ø gunuva maca na drano
non-focus
anaphoric actor AOV GOAL TOPIC
one the cloud white which drinks-it dry the pond

That is to say gunuva, an ACTOR-ORIENTED VERB, *with* a suffix, is in GOAL FOCUS, i.e. oriented towards a NP in the role of GOAL. The non-focus actor NP is Ø for the same reason as in the previous example.

... sā maca na kena wai // ni Ø sā gunuva na vara
ANAPHORIC
GOAL TOPIC AOV non-focus actor
is dry the its water because has drunk-it the pith

(The last two examples, with a reversal of relations, are thus better accounted for in terms of topic-and-focus than of subject and object.)

FIRST PROVERBIAL EXPRESSION: 'As soon as the chiefs assemble, I shall be chewed' (famous words, said to have been uttered by the kava (yaqona) shrub) (Bulicokocoko 1957:28).

Era soqo gā na tūraga // au qai mama
GOAL TOPIC GOV
They assemble just the chief I(me) then chew

That is to say mama, a GOAL-ORIENTED VERB, *without* a suffix, is in GOAL FOCUS, i.e. oriented towards a NP (in this case a preverbal pronoun) in the role of GOAL. There is no non-focus actor NP.

SECOND PROVERBIAL EXPRESSION: 'The field has been raised' (i.e. 'The home team has been beaten'). For instance: 'The Suva team have raised the field' (Anon. n.d.:30,no.32).

Sā lave na rārā // (Sā druka na itaukei)
GOV GOAL TOPIC
is lift the field (is defeat the homelander)

That is to say, lave, a GOAL-ORIENTED VERB, *without* a suffix, is in GOAL FOCUS, i.e. oriented towards a NP in the role of GOAL. As in the previous example, there is no non-focus actor NP.

Era sā mai laveta na rārā / na mata qito mai Suva
(ACTOR GOV non-focus goal ACTOR TOPIC
TOPIC)
They is come and lift-it the field / the team sport from Suva

That is to say, laveta, a GOAL-ORIENTED VERB, *with* a suffix, is in ACTOR FOCUS, i.e. oriented towards a NP in the role of ACTOR. The topic consists of two discontinuous NPs: a preverbal pronoun in the third person plural (Era) and a postverbal NP na mata qito (mai Suva).

THIRD RIDDLE: 'Two men forever fighting. One of them gets the upper hand for a long while, but one day he falls asleep. As he lies asleep his blanket is very thick and heavy. *Then along comes his enemy whom he used to defeat*, sits on top of him and overcomes him'. Answer: 'A man and grass. In his lifetime, he keeps down the weeds in his garden, but when he dies and lies under the earth, the weeds grow on top of him' (Biggs 1948:345,no.24).

Sā	qai	lako mai	na	kena	meca	/	ka	Ø	rawai	koya	e	liu
			non-focus					ACTOR	GOV			
			anaphoric goal					TOPIC				
is	*then*	*come*	*the his*		*enemy*		*who*		*defeat*	*him*		*before*

That is to say, rawai, a GOAL-ORIENTED VERB, *with* a suffix, is in ACTOR FOCUS, i.e. oriented towards a NP in the role of ACTOR. This is a particularly interesting example: the anaphora relates to a 'mooted' actor last mentioned three sentences previously (i.e. 'one of them gets the upper hand'). It is very difficult to see how one could decide, on grounds of case relations within the sentence alone (i.e. without recourse to lengthy discourse analysis) that na kena meca was in fact *not* the 'subject' of rawai. However, because a GOV *with* a suffix is oriented, as I hope to have shown, towards a NP in the role of ACTOR and although this may seem startling, even a zero ACTOR topic, as in this case, there is an adequate safeguard against ambiguity.[35]

THIRD PROVERBIAL EXPRESSION: 'The pig was killed because his legs carried him' (i.e. if a man gets into trouble in another village and is beaten up, he should not feel sorry for himself, he had no business to go there in the first place) (Bulicokocoko 1957:34).

Moku	na	vuaka	/	ni	kauta	na	yavana
	non-focus				GOV	ACTOR	TOPIC
	anaphoric goal						
kill	*the*	*pig*	*/ because*		*carry-it*	*the*	*leg-his*

That is to say, kauta, a GOAL-ORIENTED VERB, *with* a suffix, is in ACTOR FOCUS, i.e. oriented towards a NP in the role of ACTOR. Here also case grammar and IE sentence-based relations tempt one at first flush to translate this by 'The pig was beaten because he carried his legs' when in fact the opposite is intended. There seems to be no alternative explanation to considering the focal orientation of the verb (in this case a GOV with a suffix in ACTOR focus) to be the deciding factor.[36]

In conclusion let me (at the risk of claiming to remember more school Latin than I could justify) quote a sentence attributed to William of Occam, an English Franciscan friar of the 14th century, born in Ockham in Surrey and a famous theologian of his day: *'Non prodest fieri per plura quod potest fieri per pauciora'*, i.e. 'There is no advantage in achieving with more categories what can be achieved with fewer'.

NOTES

1. Yet in a review published in the same year (Schachter 1977b) he states that "Philippine languages should perhaps be classified as CASE-PROMINENT LANGUAGES: i.e. languages whose structure favors a description in which a major role is assigned to case relations" (p.710).
2. As the references given in two earlier articles show (Milner 1962; 1979:2 and 14, note 2), interest in these problems of AN grammar, particularly among Dutch linguists, goes back at least to the end of the 19th century (see also Tchekhoff 1978, especially her reference to Dirr (p.198) and

Schuchardt (p.202)). As has been pointed out before (Martinet 1965) a number of Western European languages are almost incapable of predicating anything without at least a token or dummy subject. For instance in English when talking about the weather: 'It is raining' or about abstract matters: 'It occurs to me that'. Latin however is not so dependent on formal predication as in ablative absolute constructions: *mutatis mutandis* 'if the necessary changes are made' or with infinitives: *laborare est orare; humanum est errare* 'to work is to pray; to err is human'. Martinet also shows that idiomatic and colloquial constructions such as *Y a* in French and *There's* in English cannot be considered to incorporate a subject in the strict sense of the term.

3. What he calls the 'A verbs' follow Pattern 2 only, whether they are suffixed or unsuffixed, that is to say they have an unmarked NP and an NP marked by e or 'e. 'B verbs' on the other hand:

   (i) *without a suffix*: follow Pattern 1 (an unmarked NP and an NP marked by i, 'i or ki
   (ii) *with a suffix*: follow Pattern 2 (Clark 1973:569; 574-575).

   In his doctoral dissertation Foley (1976) points out the correspondence between Polynesian A and B verbs on the one hand and Fijian *stative* and *active* verbs (to use the terms suggested by Schütz (1981:201)), respectively. Biggs (1974:424) describes the same categories in Fijian as goal-subject selecting and actor-subject selecting verbs respectively. In the present article I shall use the terms 'actor-oriented' and 'goal-oriented' verbs, respectively.

4. Pawley states that this suggestion was originally made to him by Schütz (1973:180,note 22).

5. Paraphrasing this view in her own words, Naylor writes as follows:

   > Topic and focus (in its highlighting function) belong to the system by which the clause is organized as a message .... Focus as the indicator of the participant role of the topic is at the same time a function in the system of transitivity - the organization of the clause as expression of extra-lingual reality. Emphasis, as a way of rendering something especially significant, with emotional overtones, is analyzable within the framework of the unmarked-marked distinction, which cuts across both systems. (1975:17)

6. Albert Schütz informs me (in a private letter) that David Arms' dissertation was the first full analysis of these phenomena to be published, but Bruce Biggs was already discussing consonant restrictions in the early 1960s. Paul Geraghty (1973) wrote an article on this subject and Peter Lincoln also studied the same problem.

7. For instance -c- is a replacement consonant for -v- in mawac- (*steam +*) *hit, spread to*, lawac- *start weaving (mat)*, kālawac- *step, stride over*, but not in liwav- *blow on* or dewav- (of disease) *spread to, infect* (see also note 9 below).

8. Arms did in fact consider such cases, for instance on pp.151-154 and especially in the note to p.152, but he seems to see a conflict between phonotactic rules and the semantic factors which he associates with individual thematic consonants. I see no conflict, since the phonotactic rules appear to operate rigorously, though not always clearly, since more than one rule may be involved within the same base (see in particular note 21 below).

The term *heterorganic* has been suggested to me by my colleague Eugénie Henderson who has used it in her own work to describe analogous phenomena in Khasi (Munda).

9. Following Geraghty (1983:260-270), I shall refer to *short* (i.e. monosyllabic) and to *long* (i.e. disyllabic) suffixes. In *short* suffixes:

-l- does not occur.
-d- and -t- dissociate, except in (vaka)dinat- *bear out, confirm.*
-r- and -n- dissociate, except in karon- *greatly value, take great care of.* (NB: Paul Geraghty informs me that karon- is probably cognate with qāraun- and that the restrictions may not be so strict at the distance of two vowels.)

Note also that -w- and -c- are two 'grey areas', i.e. peripheral cases where the evidence is conflicting. (For -w- see note 7 above.) -c- in some cases does not dissociate: cat- *disliking, hating*, colat- *carry on the shoulder*, cukit- *dig up the ground.* In other cases there is some evidence that -c- and -t- do dissociate. For instance, -v- in cakav- *do, make* may be a replacement for -t- because of c- in the base, see also note 21 below.

In *long* suffixes:

-g- does not occur.
-l- and -r- occur in suffixes which have specialised functions ('repetition' and 'intensity, force or violence' respectively).

10. I have consulted the dictionaries of Hazlewood, an unpublished dictionary by J. Neyret (which exists in typescript form and is available in the library of the National Archives of Fiji as well as in the library of the School of Oriental and African Studies, University of London). I am also indebted to Tevita Nawadra, Director of the Fijian Dictionary Project, for his permission to consult his files between July 1980 and March 1981.

11. Naylor (1975:21-22) uses the term to describe the 'locative' focus -an, i.e. "location in physical and psychological space".

12. When, as happens in a number of cases, both occur after the 'same' base, there is a difference of lexical range of application which sometimes suggests that two homophonous bases are represented: e.g. bulia *form, shape*; bulita *adorn (a canoe) with white shells* (buli). qalia *roll, twist (sinnet) on thigh*; qalita *snatch a corpse after a battle.* (NB: bulita and qalita appear in Hazlewood's dictionary.)

13. For instance:
sokot- *sail (in) a certain wind or weather*
sokov- *sail through or towards*
rokot- *bend (bow or stick)*
rokov- *bow to, pay respect to*
kosot- *cut something with shell or knife*
kosov- *cut lengthwise; cut across*

14. For instance:
Replacement of -v- by -c-: vukac- *fly towards*; kuvuc- *(of smoke) puff against*
kawac- *go over, cross (a bridge)*

Replacement of -v- by -k-: virik- *throw something at*; dumuk- *push upwards (with a pole)*;
yawak- *get away from*

Replacement of -v- by -r-: vocer- *paddle to (a place)*; caber- *carry something up*;
cabor- *offer up*

Replacement of -v- by -t-: kevut- *climb down along or towards*; kabat- *climb up to*;
robot- *extend over*

15. For instance:
Replacement of -t- by -c-: tukuc- *lower (something suspended)*; taloc- *pour carefully*;
talac- *remove, shift*

Replacement of -t- by -k-: dirik- *smash (shell), knock* (tabua); natuk- *knead*;
tonok- *poke with finger +*

Replacement of -t- by -r-: taqar- *lay, place (on top of)*; utur- *place end to end*

Replacement of -t- by -v-: dolav- *open*; takiv- *draw water*;
kotiv- *cut, clip (hair, paper +)*

16. e.g. tarog- *ask (a question)*; bilig- *push*; vivig- *roll (a mat, etc.)*. There is some evidence to suggest that in some cases 'semantic analogy' may be one of the factors involved. For instance olog- *wrap in a bundle*; salag- *wrap in leaves for cooking*.

17. e.g. sodom- *insert, fit (s.th. cylindrical)*; daram- *slip (into ring or sheath)*.

18. e.g. curum- *go through, go in (or out)*; nanum- *keep in mind, think of*; sucum- *give birth to, suckle*.

19. e.g. tawan- *settle in, occupy*; tokon- *prop up*; yaben- *lead, support (old or sick person when walking)*.

20. e.g. qāraun- *look after, take care, beware*; kumun- *store up, collect*; tukun- *tell*.

21. For instance:
(i) -t- may be replaced by -c- in: dabec- *sit on*; davoc- *lie on*, and that
(ii) -v- may be replaced by -r- in: (vaka)daber- *set down, place in sitting position*; (vaka)davor- *place (child, etc.) in lying position*
but the evidence is conflicting. Thus in cakav- *make, do*, -t- may be replaced by -v-. It is also important to note that in some cases, the occurrence of replacement consonants which cannot be accounted for by the phonotactic rules given above, can be explained by diachronic factors. That is to say, in some favourable positions a PAN consonant may have been retained owing to the fact that its modern reflex coincides with an acceptable replacement consonant. See for instance tagic- *cry for* (i.e. to obtain something) (Milner 1986).

22. Each of them in long suffixes is followed by -aka (or -aki). It should be noted that *-gaka (or *-gaki) does not occur, but -laka (-laki) does, although -l- does not occur as a thematic consonant in short suffixes. (NB: -g-, which does not occur in long suffixes, does so in short.)

23. e.g. tagotaka *borrow temporarily*, qitotaka *play with (ball, etc.)*, ridotaka *hop with something*. In addition to its use with bases that can also be followed by short suffixes, -taka (-taki) is frequently used to derive denominal verbs.

    It is especially interesting to note that although -t- in a long suffix (-taka or -taki) does not dissociate from t or d in the preceding base, -vaka (or -vaki) occur in a number of cases where -taka (or -taki) might have been expected to occur for reasons which are discussed in the next two paragraphs and illustrated in notes 24 to 28. A possible reason is that the *long* suffix after certain bases is formed by analogy with the thematic consonant in the *short* suffix. For instance:
    ladevak- *jump with something* (by analogy with ladev- *jump over*)
    talevak- *return s.th. borrowed* (by analogy with talev- *go again to*)
    takivak- *scoop, ladle (with s.th.)* (by analogy with takiv- *draw (water)*
    (See also note 31 below.)

24. For instance:
    instrument: viritak- *throw*; cokatak- *hurl*; nimatak- *use as a bailer.*

25. *object affected:* drōtak- *run away with*; kabatak- *climb (carrying s.th.)*; karatak- *propel, punt (a canoe, etc.).*

26. *object produced:* kāsivitak- *spit (saliva, etc.)*, vekacak- *pass (a stool), excrete (faeces).* (NB: -c- replacing -t- by analogy with vekac-; see note 31.)

27. For instance:
    on account of, about: rogovak- *spread report on account of*; dredrevak- *laugh about (or over)*; surevak- *beg earnestly, entreat for.*

28. *on behalf of, for the benefit of:* serevak- *sing about*; osovak- *bark because of*; drōvak- *run away because of.*

29. e.g. seretak- *sing (a song)*
    serevak- *sing (about s.th.)*
    tagitak- *utter (a cry, etc.)*
    tagivak- *cry on account of s.th., lament*

30. e.g. valatak- *fight for s.o. (or s.th.)*; vūnautak- *preach on (a topic)*; meketak- *sing and dance on account of.*

31. Thus cicivak- seems to be formed by analogy with ciciv-, likewise kilicak- (kilic-), mīcak- (mic-), curumak- (curum-), lakovak- (lakov-).

32. For example -kak- occurs in rukak- *curse* and tavukak- *singe (pig).* -rak- occurs in wārak- *wait for* without any association with force, violence or intensity.

33. cōnak- *cover floor with grass and/or mats*; savenak- *hang (sail) by the* save (save is the name of a rope hanging from the mast-head).

34. He goes on to say that Fijian in this respect "is reminiscent of Polynesian languages such as Futunan and perhaps Tongan and Samoan" but "quite unlike English where ... all verbs are actor-subject selecting, until they are passivised" (Biggs 1974:418). One could argue, however, that English

reflects an analogous distinction, at least covertly, if not overtly. Thus a door may 'open' or 'shut' but one can hardly say that *'milk drinks' or that *'bread eats'. Likewise in the imperative, 'eat' and 'drink' can be used without an overt object, but 'open' and 'shut' cannot be so used, except perhaps by a dentist.

35. In order to make absolutely sure that there was no ambiguity, I asked Tevita Nawadra, the Director of the Fijian Dictionary Project, what he would have said if, in fact, it had been 'his enemy who used to defeat him'. His answer was: Sā qai lako mai na kena meca, ā dau rawai koya e liu. Here rawai is also in ACTOR FOCUS (as indicated by the suffix) but na kena meca is the anaphoric ACTOR TOPIC.

36. According to Tevita Nawadra, it would also be possible here to have na yavana in focus in the role of GOAL. Since, however, we are dealing with a GOAL-ORIENTED verb, if it was oriented towards a GOAL TOPIC it would, by definition, not have a suffix. The only possibility, therefore, would be: Moku na vuaka, e kau na yavana *The pig has been killed, its legs have been carried away* (or: *someone has carried away its legs*). The implication would then be that a pig had been slaughtered, cut up and its legs taken somewhere else for whatever reason (distribution, roasting, etc.).

POSTSCRIPT

I now have the pleasant duty of thanking the Government of Fiji for their generous assistance, which made it possible for me to attend the FOCAL conference and to present a paper on which this article is based. I am also indebted to Paul Geraghty for his painstaking and constructive criticism, as well as to Otto Chr. Dahl, Eugénie Henderson, Tevita Nawadra, Paz Naylor, Vula and Chris Saumaiwai and Albert Schütz.

REFERENCES

ANON.

n.d. Idioms (list prepared for use in Fijian schools). TS.

ARMS, D.G.

1973 Whence the Fijian transitive endings? *Oceanic Linguistics* 12:503-558.

1975 Transitivity in Standard Fijian. Ph.D. thesis, University of Michigan, 1974. Ann Arbor, Michigan: University Microfilms International.

BIGGS, B.G.

1948 Fijian riddles. *Journal of the Polynesian Society* 57:342-348.

1974 Some problems of Polynesian grammar. *Journal of the Polynesian Society* 83:401-426.

BULICOKOCOKO, S.T.U.

c.1957 Fijian idioms, expressions, proverbs and colloquialisms. TS.

CAPELL, A.

1941 *A new Fijian dictionary.* Sydney: Australasian Medical Publishing Company.

CHURCHWARD, C.M.

1941 *A new Fijian grammar.* Sydney: Australasian Medical Publishing Company.

CLARK, R.

1973 Transitivity and case in Eastern Oceanic languages. *Oceanic Linguistics* 12:559-605.

1977 Eastern Oceanic transitive suffixes and the genesis of rules. Paper presented to the Austronesian Symposium, L.S.A. Summer Institute, University of Hawaii.

DAHL, O.C.

1978 The fourth focus. In Wurm and Carrington, eds 1978:383-393.

FERRELL, R. and P. STANLEY

1979 Austronesian versus Indo-European: the case against case. In Naylor, ed. 1979:19-31.

FIRTH, J.R.

1948 Sounds and prosodies. *Transactions of the Philological Society* 127-152.

FOLEY, W.A.

1976 Comparative syntax in Austronesian. Ph.D. thesis, University of California, Berkeley.

GERAGHTY, P.

1973 Some aspects of case-marking in Fijian. Term paper, University of Hawaii.

1983 *The history of the Fijian languages.* Oceanic Linguistics Special Publication No.19. Honolulu: University of Hawaii Press.

HAZLEWOOD, D.

1872 *A Fijian & English and an English & Fijian dictionary ... and a grammar of the language ...* edited by J. Calvert. 2nd edn; 1st edn 1850. London: Sampson Low, Marston.

HOCKETT, C.F.

1976 The reconstruction of Proto Central Pacific. *Anthropological Linguistics* 18:187-235.

1977 Proto Central Pacific: addenda. *Anthropological Linguistics* 19: 242-250.

HOPPER, P.J. and S.A. THOMPSON

1980 Transitivity in grammar and discourse. *Language* 56:251-299.

KRUPA, V.

1966 The phonemic structure of bi-vocalic morphemic forms in Oceanic languages. *Journal of the Polynesian Society* 75:458-497.

LICHTENBERK, F.

1978 Thematic consonants in Manam transitive verbs. *Anthropological Linguistics* 20:185-193.

MARTINET, A.

1965 La construction ergative. In: *La linguistique synchronique*, 206-222. Paris: Presses Universitaires de France.

MILNER, G.B.

1962 Active, passive, or perfective in Samoan: a fresh appraisal of the problem. *Journal of the Polynesian Society* 81:5-18.

1972 *Fijian grammar*. 3rd edn. Suva: Government Press.

1979 On the centrality of Austronesian syntax. In Naylor, ed. 1979:1-17.

1986 On prosodic relations between Fijian bases and verbal suffixes. In J.H.C.S. Davidson, ed. *Contributions to South East Asian linguistics: Essays in honour of Eugénie Henderson*. London: School of Oriental and African Studies. (In the press.)

NAYLOR, P.B.

1975 Topic, focus, and emphasis in the Tagalog verbal clause. *Oceanic Linguistics* 12-79.

1978 Toward focus in Austronesian. In Wurm and Carrington, eds 1978: 395-442.

1979 Linking, relation-making, and Tagalog syntax. In Naylor, ed. 1979: 33-50.

NAYLOR, P.B., ed.

1979 *Austronesian studies: papers from the Second Eastern Conference on Austronesian Languages*. (Michigan Papers on South and Southeast Asia 15, 1979). Ann Arbor, Michigan: University of Michigan.

NEYRET, J.

1935 Fijian-English dictionary. TS.

PAWLEY, A.

1973 Some problems in Proto-Oceanic grammar. *Oceanic Linguistics* 12: 103-188.

PIKE, K.L.

1963 A syntactic paradigm. *Language* 39:216-230.

SCHACHTER, P.

1976 The subject in Philippine languages: Topic, actor, actor-topic, or none of the above? In C.N. Li, ed. *Subject and topic*, 493-518. New York, Academic Press.

1977a Reference-related and role-related properties of subjects. In P. Cole and J.M. Sadock, eds *Syntax and semantics*, vol.8: *Grammatical relations*, 279-306. New York: Academic Press.

1977b Review of Ramos, T.V. *The case system of Tagalog verbs*. *Language* 53:707-711.

SCHACHTER, P. and F.T. OTANES

1972 *Tagalog reference grammar*. Berkeley: University of California Press.

SCHÜTZ, A.J.

1981 Specification as a grammatical category in Fijian. In J. Hollyman and A. Pawley, eds *Studies in Pacific languages and cultures: In honour of Bruce Biggs*, 197-207. Auckland: Linguistic Society of New Zealand.

1986 *The Fijian language*. Honolulu: University of Hawaii Press. (In press.)

SCHÜTZ, A.J. and T.R. NAWADRA

1972 A refutation of the notion "passive" in Fijian. *Oceanic Linguistics* 11:88-109.

SCOTT, N.C.

1948 A study in the phonetics of Fijian. *Bulletin of the School of Oriental and African Studies* 12:737-752.

TCHEKHOFF, C.

1973a Some verbal patterns in Tongan. *Journal of the Polynesian Society* 82:281-292.

1973b Verbal aspects in an ergative construction: an example in Tongan. *Oceanic Linguistics* 12:607-620.

1978 *Aux fondements de la syntaxe: l'ergatif*. Paris: Presses Universitaires de France.

WOLFF, J.U.

1979 Verbal morphology and verbal sentences in Proto-Austronesian. In Naylor, ed. 1979:153-167.

WURM, S.A. and Lois CARRINGTON, eds

1978 *Second International Conference on Austronesian Linguistics: proceedings*. *PL*, C-61.

FOCUS IN MALAGASY AND PROTO-AUSTRONESIAN

Otto Chr. Dahl

1. INTRODUCTION

1.1 The Malagasy verb has a focus system very similar to that found in many Philippine, Minahasan and Formosan languages. We shall here study in detail the morphology and syntax of the Malagasy system, comparing it with the grammar of other languages and with what may be supposed to be Proto-Austronesian.

1.1.1 The phonetic development and the vocabulary of Malagasy are so similar to the languages of the SE Barito subgroup in Kalimantan that it undoubtedly belongs to this subgroup (see Dahl 1977). The ancestors of the Malagasys seem to have migrated to Madagascar about 400 A.D. (Dahl 1951:366-369).

1.1.2 The only Bornean language of this subgroup from which more than wordlists has been published, is Ma'anyan. In this language there is some literature available (see Dahl 1951:24-25) and an outline of a grammar (Sundermann 1913). From this grammar and the texts it is clear, however, that Ma'anyan does not possess the focus structure. It has active forms with affixes that are recognisable in Malagasy, and a passive form less easily comparable.

But the distance between Madagascar and the northern islands of western Austronesia is so great that a separate development of a complicated system with nearly identical forms in each of these widely separated areas must be regarded as impossible. We are therefore forced to assume that SE Barito had the focus system at the time of the emigration towards Madagascar, and that these languages have lost it during the intervening 1600 years.

Such changes are by no means surprising. The Romance and most Germanic languages have lost the old Indo-European case system in the same or even shorter time than is assumed here. The case system is still present in southern German and in Icelandic, that is to say on the fringes of the Germanic area. It should accordingly come as no surprise to find the focus system in the periphery of Austronesia, since this merely illustrates the general tendencies of language families to develop more rapidly in central areas than in their more conservative fringes.

1.1.3 Since phonetically conservative Formosan languages like Paiwan possess the focus system, it is reasonable to assume that this system belonged to PAN grammar.

1.2.1 However, before studying the Malagasy focus system in detail it is necessary to consider some important features in the phonetic development of the language after its arrival in Madagascar. The Barito languages have both consonants and vowels in final position like so many other AN languages. Indeed,

---

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 21-42. *Pacific Linguistics*, C-93, 1986.

this characteristic is so widespread that it is believed to have existed in PAN. In Malagasy all words have only vocalic finals. Other phonetic changes, similar to the phonetic development of neighbouring Bantu languages, indicate that the immigrants found and absorbed a Bantu population in Madagascar. This Bantu substratum then influenced the Austronesian language of the colonists. The Bantu languages of East Africa have only vocalic finals, and the change is supposed to have occurred because the substratum found it difficult to pronounce final consonants.

In all dialects -l, -s and -h have been apocoped. To -k, -t and -r a final vowel has been added, but -t and -r have merged into an affricate. In Merina the result is -tra, in Sakalava -tse. Final -p has mostly merged with -t, less frequently with -k. These changes are only word finally.

In Merina the final nasals have all merged into -n, which has added a final -a like -ka and -tra. In Sakalava the final nasals have been apocoped, like -l, -s and -h (see Dahl 1954, especially pp.343-344).

When a suffix with initial vowel is added, no changes in the wordbase are required by the structure of the substratum. Before such suffixes the final consonant of the wordbase is therefore often maintained in the shape it now has in intervocalic position. But before the suffixes -ko '1st pers.sg.' and -ny '3rd pers.' the n of -na is also deleted.

1.2.2 Ma'anyan has a non-phonemic penultimate stress. When a final vowel was added in Malagasy, the number of syllables increased by one. But the accent has remained on the syllable that was penult before the lengthening. In words ending in -ka, -tra or -na in Merina it now therefore falls on the antepenult. Like vowels that have come into contact by loss of a consonant, have been contracted. The same has occurred when a suffix with initial vowel has been affixed to a word ending in the same vowel. This reduces the number of syllables. But the accent remains on what was the penultimate vowel before the contraction. Therefore the accent may now also fall on the ultimate syllable.

As a result of these changes Malagasy places a phonemic accent on one of the last three syllables of the word, e.g. Mer[1] tánana *hand, arm* (< PMP *taŋan id.), tanána *village, town* (< *tanáan < PMP *tanaq *land* + *-an), manála *to take away*, manalá *take away!* (imper. < *ma-n-alá-a < PAN *ma-ŋ-ala + -a).

2. ACTOR FOCUS

As a rule languages with focus structure have four different focus forms generally called actor focus (AF), object focus (OF), referent focus (RF), and instrument focus (IF). All these are found in Malagasy.

2.1 Actor focus, which has the performer of the action in focus (mainly as subject), is often formed with the infix -um- in the Philippines and Formosa. In Malagasy dialects there are still some very few cases of this. In Sakalava we have l-om-aŋo[2] *to swim* < PAN *ɬ-um-aŋui, t-om-áŋy *to cry* < PAN *t_2-um-aŋit, h-om-éhe *to laugh*, cf. Old Javanese kəkəl id., l-om-áy *to run*.[3] These forms are present tense. In the past tense the infix is replaced by the prefix no-: no-láŋo, no-táŋy, no-héhe, no-láy.

2.2 In Sakalava we also have h-óm-a *to eat* < PAN *k-um-aʔən. This word is present in Merina too, in the form h-óm-ana, with the regular development of

final n into na. In this word the crasis of a + ə has already taken place in Kalimantan, cf. Mny kuman id. The word has thus come to Madagascar in this form, with the stress on the ú. The fact that the wordbase was so short explains the abnormal accentuation of the infix. The word has consequently not been regarded as a form with infix, but rather as a wordbase used as verb with no separate form for the past tense.

In Merina, homana is the only word of this shape used as a verb. The first three forms above are also considered as wordbases in Merina, and form AF with the prefix mi-: mi-lománo *to swim*, mi-tomány *to cry*, mi-homéhy *to laugh*. So does the fourth one, but this has been transformed by popular etymology into mi-olo-may *to run with haste*, lit. *to act as a burning person*.

2.3 The regular formation of AF in Malagasy is with the prefixes ma- + nasal substitution or accretion, generally transitive, and mi- (probably < PAN *maɣ-), often intransitive. Besides lomáŋo *to swim* (intr.) mentioned above, Sakalava has man-dáŋo *to cross swimming* (tr.). Other examples from Merina: mamábo or mam-bábo *to capture, seize as prey*, mamérina *to send back* (tr.), mi-vérina *to come back* (intr.). A few wordbases form AF with ma- without nasal substitution or accretion, e.g. Mlg ma-hita *to see* < PAN *kit$_1$a? id. Similar verbal forms are found in Philippine languages.

2.4 In the past tense the initial m of all these prefixes is replaced by n, and in the future tense by h, e.g. na-hita *saw*, ha-hita *will see*. The origin of the n is probably that the prefix has got the infix -in-: m-in-a-, as seen in some AN languages. Thereafter the initial syllable has undergone aphesis: m-in-a- > na-. In other languages this -in- more often seems to indicate perfective aspect than past tense. But a relation between perfective and past is quite understandable.

The origin of no- in the past tense of the above-mentioned Sakalava verbs is probably the same. In some AN languages there is a prefix mu/mo- with the same function as the infix -um-. I suppose that the prefix has been the original form, and that the infix has come into being by metathesis with the initial consonant of the wordbase: mu-C ... > C-um- .... Sak no- is then *m-in-u- > nu- following the same process of development as above.[4] We shall see that in Malagasy this no- has developed a broader function as formative of the past tense, even in forms that do not contain -um-.

2.5 To the h- of the future tense I have not found any parallel in AN languages. In other verbal forms the morpheme of the future is ho. It is possible that this is due to the Bantu substratum. In neighbouring Bantu languages a prefix ku/hu- is part of the future morpheme. The substratum, which had a verbal system with tenses, may have felt the necessity of a future tense in the verb, which the language of the AN immigrants lacked (see Dahl 1954:355-360), and it has also transformed the AN perfective aspect into a past tense.

3. OBJECT FOCUS

3.1 The morphemes of object focus in Philippine and Formosan languages are reflexes of PAN *-ən. In Malagasy too we find the same with regular phonetic reflexes: Mer -ina, Sak -e. In Sakalava the reflex of PAN *ə is e in all positions. In Merina it is e in the accentuated syllable and in the preceding ones,

but in the syllables following the accent it is i. Since PAN *-ən was in the ultima, it was not accentuated in Barito, and therefore neither in Malagasy, e.g.:

(1) tehén -ina ny lákana
push forward OF *the canoe*
the canoe is pushed forward (with a staff)

< PAN *t_2əkəɬ + -ən, cf. Mer téhina *staff, stick* (cf. 1.2.1 and 1.2.2 above).

3.2 But if the wordbase had final e in Proto-Malagasy the two vowels in contact are contracted to an accentuated é, e.g. Sak teré *to be milked*, Mer teréna id., *to be pressed*, from *teré-en < PMP *təd$_1$ət' + -ən (with apocope of a final s). If, however, the wordbase had final i in Proto-Malagasy, Merina shows contraction into í, e.g. Mer fidína but Sakalava filí-e *to be chosen* from PAN *piliq + -ən (apocope of q already in Barito). The rule of contraction of like vowels has thus had its effect both before and after the change of PMlg *e > Mer i in syllables following the accent.

3.3 In the future tense all dialects have ho- before consonant and h- before vowel, cf. 2.5 above. In the past tense Merina has no/n- in harmony with this. But in Sakalava we often find in the past tense the same as in Philippine languages: the suffix is omitted, and instead we have the infix -in-, e.g. to filí-e *is chosen* we have f-in-íly *was chosen*, and to laŋós-e *to be swum in, across* we have l-in-áŋo *was swum in, across*. This correlation between -ən and -in- seems to be old in AN. But it is also possible in Sakalava to prefix ni- before the whole form with -e, e.g. ni-filí-e *was chosen*.

3.4 This form has in focus the object suffering the action, and this is the subject of the clause, see example (1) above.

4. REFERENT FOCUS

4.1 The morpheme of referent focus in Philippine and Formosan languages is -an in almost all languages where it occurs, and this has also been supposed to be its PAN form. Starosta, Pawley and Reid have, however, reconstructed it as PAN *-ana, based on Oceanic, Malagasy and Tsou (1982a:163, 1982b:104).

4.1.1 It is correct that -ana is the morpheme of RF in Merina and some other Malagasy dialects. But the final a in this suffix is an innovation in Malagasy due to the transition from consonantal to vocalic finals, see 1.2.1 above. That this is so appears clearly in Sakalava and some other dialects, where the RF suffix is -a. Here Ø is the regular reflex of *final* n. The development of -ana/a from PMlg *-an is exactly the same as the development of OF -ina/e from PMlg *-ən.

If the suffix had been **-ana in Barito, the penultimate a would have been accentuated, and would have continued to be stressed in Malagasy. But Mer -ana is unaccentuated, except when there is contraction with a final accentuated a in the preceding syllable of the wordbase, see 1.2.2 above, e.g. sorát-ana *is written (on)* < PAN *t'urat$_1$ + -an, alána *is taken away* < alá-ana < PAN *ala + -an.[5]

Mer -ana can therefore not be used as an argument for PAN **-ana. Comparative linguistics is diachronic linguistics, and a merely synchronic comparison without attention to the historical background may be misleading.

Ma'anyan, which now has no focus forms, has a suffix -an, never **-ana, that is used in derivative forms with different meanings. And as far as I know no other languages in Western Austronesian have the suffix in the form -ana. There is therefore no valid argument from Western Austronesian for a PAN **-ana.

4.1.2 The three authors' argument concerning Tsou depends on Tung 1964:174-175. Under the heading "derivational suffixes" Tung writes:

> /-ana/, attached to certain conjoined words (being place and clan names in meaning), is very much like the English suffixes '-place', '-town', '-man', 'smith' and so on in function.

The combination with clan names that may also be derived from names meaning human professions, raises the question as to whether the function of the suffix is really locative. At any rate, it is not a morpheme of RF. In Tsuchida 1976: 102-103 we find that "the location focus marker is -i".

Tsou has, like Kanakanabu and Saaroa, vocalic finals. Tsuchida says (p.88): "In word final position a morphophonemic form ending in a consonant or stressed vowel is realized with a supporting vowel".

We cannot therefore discount the possibility that Tsou -ana may have developed from *-an. However, as supporting vowels Tsuchida cites i, u and ə, but not a in modern Tsou. A development -ana < *-an is thus not certain. However, on balance, the argument for PAN *-ana from the derivational suffix -ana is far from convincing.

4.1.3 Starosta, Pawley and Reid do not give any details about -ana in Oceanic, either where it occurs or its function, and I have not had the opportunity to study it closely. If it does occur sufficiently often there, *-ana may be constructed as a Proto-Oceanic innovation, but not as PAN. From the very frequent occurrence of the form -an of the RF suffix in Formosa, the Philippines, and in Proto-Malagasy I consider *-an to be the most likely PAN form.

4.2 Before studying the use of -ana/a in Malagasy we have to note a phonetic feature in Merina. If the final vowel of the wordbase is or has been e, Merina has crasis of e + a into é, but Sakalava has generally not, e.g. omé[6] *to give* + -a(na) is Mer oména, but Sak omé-a *to be given.*

4.2.1 What is focused with the form with -ana/a in Malagasy is not so uniform as with -ina/e. It may be the place where the action is located, e.g.

(2) Mer totóf-ana tány ny lávaka
fill RF *earth the hole*
The hole (in the ground) is being filled with earth.

Here the direct object of the action is the earth, tany is constructed as such, and the hole is the location of the action and subject of the clause.

4.2.2 In other cases the person profiting from the action, he who receives the direct object of the action, is in focus and constructed as subject, e.g.

(3) Mer tolór -ana fanomézana ny vahíny
hand over RF *gift the guest*
The guest is presented with a gift.

The gift is the direct object. In AF both these verbs may be constructed with two objects:

(2a) Mer manótotra tány ny lávaka ízy
He fills the hole with earth. OR *He fills earth into the hole.*

(3a) Mer manólotra fanomézana ny vahíny ízy
She gives the guest a gift, or

(3b) Mer manólotra fanomézana ho an'ny vahíny ízy
She presents a gift to the guest, with one object and one complement.

4.3.3 But in many cases a direct object seems to be in focus, e.g.

(4) Mer mamáfa tokotány ízy
AF *sweep courtyard he*
He sweeps the courtyard.

(4b) Mer fafána ny tokotány
sweep RF *the courtyard*
The courtyard is being swept.

An explanation for the apparent anomaly may be that in this case the courtyard is both object and location of the action. And there are other verbs that allow the same interpretation, e.g.

(5) Mer sorát-ana ny taratásy
write RF *the paper*
there is written on the paper or *the letter is written*

But in

(6) Mer sorát-ana ny téni-ny
write RF *the word his*
his words are written down; only a direct object is in focus.

4.3.4 AF of man-omé can take two objects, and both may be focused by RF oména:

(7) man-omé ny vahíny ny sakáfo ianáo
AF *give the guest(s) the food you*
You give the guest(s) the food.

(7b) Mer omén -áo ny vahíny ny sakáfo
give RF *you the guest(s) the food*
The food is given (to) the guest(s) by you. Or

(8) Mer manomé sakáfo ny vahíny ianao
You give the guest(s) food.

(8b) Mer omén-áo sakáfo ny vahíny
The guests are given food by you.

In the latter case the focus is benefactive, but in the former there is no trace of benefactive or locative.

4.3.5 Many forms with -ana/a have only the direct object in focus, e.g.

(9) arí- ana ny fakofáko
throw away the rubbish
The rubbish is thrown away.

In such cases -ana has the same function as -ina, and it is often impossible to understand why -ana is chosen instead of -ina.

4.4 In Merina forms with -ana have the same tense prefixes as -ina, past tense no/n-, future tense ho/h-. In Sakalava we mostly find ni- and ho- combined with -a, but never the infix -in-. However, Malagasy must earlier have used -in- with forms with -an. The name of a certain town in Betsileo is Am-bato-f-in-andráh-ana *at the stone where things have been chiselled* or *at the chiselled stone*, from fándraka *chisel*.

5. INSTRUMENT FOCUS

5.1.1 In many Philippine languages there is a form with i- focusing the instrument or the means used to perform the action. In Formosan languages the prefix is generally si-, but in Bunun is-. From these reflexes I have reconstructed the prefix as PAN *Si- (Dahl 1973/76:119).

5.1.2 However, Starosta, Pawley and Reid do not find my reconstruction sufficiently motivated, and prefer to reconstruct it as PAN *iSi- with the following motivation:

> Dahl ... reconstructs this form as *Si- for PAN, in spite of the fact that this would be expected to produce hi- in Tagalog, rather than the ?i- that is actually attested .... In Bunun, there is a similar form, but it is is- rather than si-, and marks future AF as well as IF. ... the reconstruction of *iSi- provides a better explanation of the reflexes in Bunun and Philippine languages than does *Si-. Bunun is- can be accounted for as a result of vowel loss rather than metathesis, whereas Philippine ?i- forms can be assumed to have developed by reduction of the Philippine reflex *ihi- to *?i-. Northern Philippine languages which reflect PAN *S as glottal stop (or zero) would have reduced *i?i- to *?i-. A few Philippine languages still show hi- rather than ?i- as the IF prefix.
>
> (Starosta, Pawley and Reid 1982a:165)

5.1.3 But as far as I know no language other than Bunun has the sequence is-, and no language has reflexes of all the three phonemes in **iSi-. If this was the original form, we should expect the form **ihi- in some Philippine language.

In private correspondence R.D.P. Zorc has given me the following survey of the reflexes of the prefix in Philippine languages:

> There is *no* Philippine and no other Formosan evidence for a PAN *iSi-, only *Si-. The only Philippine language that gives clearcut evidence for *Si- is Tausug with a hi- punctual instrument and a hipag- durative instrumental prefix, i.e., some form of hi- (with h < *S) is retained *throughout the grammatical system*. Samar-Leyte has mahi- and nahi- in the potential instrumental system, but simply ?i- in the punctual and durative; similarly, Alangan has an accidental instrumental prefix hi- (future, in contrast with a ha- past < PAN *Sa-), but otherwise uses ?i- as the normal instrumental prefix in the durative and punctual

systems. (See Zorc 1977:117-118, for a description of the durative vs punctual systems.) Obviously, the Aklanon and Samar-Leyte evidence is conflicting; Akl seems to indicate a split of PAN *Si- into an irregular (i.e. loss of h < *S) form normally used in the system, with the accidental form (if from the same PAN *Si-) retaining the h. Same problem with S-L.

5.1.4 I know no case of PAN intervocalic *S into Philippine ʔ. The intermediate form **iʔi- is therefore very hypothetical. In initial position, however, some Philippine languages have cases where ʔ seems to reflect PAN *S (see examples in Dahl 1981:45-46). If we assume that the proto-form of the prefix was *Si-, it thus fits better with the reflexes in Philippine than **iSi.

Zorc mentions Tausug and Aklanon as languages with an uncomposed hi-. And both in these languages and in Samar-Leyte we have hi- in composed prefixes. There are therefore reasons to believe that the Proto-Philippine form of the prefix was *hi-, and that the h was later lost in most Philippine languages. The h has the weakest articulation of all consonants, and a regular or irregular loss of it is therefore frequent in the history of languages. French orthography shows that this language has lost h twice. Malagasy is now losing it for the third time in its history: 1) PMP *h < PAN *S, 2) SEBarito h < Barito s < PAN *t', 3) and now the Merina dialect is losing h < PAN *k. That h is lost in a prefix, even irregularly, is not very surprising. A syllable at some distance before the accentuated one has often a feebler articulation, and its frequent use further weakens it. Pronounced distinctly or not it is always understood from the context.

When this initial h had been dropped, the i- was in initial position. Many Philippine languages have developed glottal onset to initial vowels, and automatically the IF i- must also be articulated in the same way. I therefore assume that the glottal stop here does not directly reflect PAN *S, but has developed secondarily as a normal part of the articulation of initial i.

5.1.5 If the original form of the IF prefix was *Si-, we have to explain how it has become is- in Bunun. Metathesis is frequent in this language, especially in the Isbukun dialect. Compare the following forms in Bunun dialects: Metathesis of consonants: Ttd, Tkb lisáv, Isb síʃav *leaf*; Ttd, Tkb qopsíl, Isb xóspil *hair*. Metathesis of vowel and consonant: Ttd, Tkb qaʃóaʔ, Isb ʔaxʃóa *ants*; PAN *$t_2$aliŋaH_2, Ttd taiŋáh, Tkb taíŋaʔ, Isb taŋía *ear* (metathesis of vowel and consonant or of ŋ with an original l that was later dropped). Metathesis of vowels occurs in all dialects: PAN *at'əŋ, Ttd isʔaŋ, Tkb isʔá:ŋ, Isb ísʔaŋ *breath*; PAN *qaS_1əlu[H_2], Ttd qosáoh, Tkb qosá:oʔ, Isb xosáo *pestle* (Tsuchida 1971:4,6,9,13, 19). With this frequent occurrence in Bunun the hypothesis of metathesis from *Si- into is- seems very reasonable.

5.1.6 In addition, the morphology of the IF prefix in Bunun gives important information about its history. In its past and perfective forms it is combined with the infix -in-, and in this form is s-in- without the initial i (Ferrell 1972:123). When in is *infixed*, its place is always *behind the initial consonant*. When it is affixed to a word with *initial vowel*, it is *prefixed*. The composed prefix should thus have been Bun **in-is- if it had initial vowel at the time of the combination of the two morphemes. The affix in is a very old morpheme in Austronesian (cf. Starosta et al. 1982a:163; 1982b:121), and the combination of the two morphemes is therefore likely to have taken place far back in history. The form s-in- reveals that the IF prefix had *initial* s when the combination

took place. In my opinion this proves that the original form of the prefix in Bunun was si-. A metathesis of it has thus taken place later, produced by the trend towards metathesis found in this language.

5.1.7 In Atayal the IF prefix is s- (Egerod 1965:269; Ferrell 1972:124).[7] In this language the vowel following the initial consonant is very often lost in the non-active focuses (Egerod 1965:255), which explains the reduction of *Si- into s-. That this *S in Atayal has the reflex s and not h, shows that the proto-form of it was PAN *S_1.

5.2.1 I have shown that the modern Malagasy morpheme of this focus is not i, which would be the normal reflex of *S_1i-, but it is a-. Moreover, this form has not only the instrument in focus, but even more frequently a moving object. With the prefix i- the same is seen in some Philippine languages. No prefix cognate to *S_1i- was found in Malagasy when I treated these prefixes, but I assumed that it had existed earlier in the history of this language (Dahl 1978, especially p.389).

5.2.2 One of the Malagasy dialects, Antemoro, has a literary tradition, at least 500 years old, written in Arabic script. No texts exist that can be proved to be so old, because they are written on a locally produced paper which is not sufficiently durable. The oldest texts have therefore been copied several times, and may have been 'modernised' by copyists. But magic texts have a more archaic language than the others. The least change in a magic text may cause the loss of its magic power, and it must therefore be copied more scrupulously. In these texts I have now found instrumental forms with i-, which do not exist in modern Malagasy, e.g.

(10) sóratsy hi- távo- ny áma hinóm-i- ny
writing IF *anoint he and drink* OF *he*
Writing with which he shall anoint himself and which he shall drink.

Here h-i-távo-ny is future tense of IF with i- and with third person pronoun suffix, and h-inom-i-ny is future tense of OF with -i(na) and the same pronoun suffix. The written magic text is supposed to be dissolved in water and is the means to be used for anointment and the object to be drunk.

However, the same texts also contain forms with the prefix a-, but these have a moving object in focus, e.g.

(11) ronóno -n' ólon a- íliŋy aŋ- óroŋy
milk gen. *human being bring down in nose*
Human milk is poured by him into his nose.

(aíliŋy is composed of a-íliŋ + -ny, and aŋ-óroŋy of a locative prefix + óroŋ + -ny, both with assimilation of the nasals) (Dahl 1983:36-38 and 1971).

5.2.3 I proposed in 1978 as a tentative hypothesis that two prefixes, one with an instrument in focus, the other with a moving object in focus, had merged and combined the two functions (Dahl 1978:389). This was a guess, because no form with i- had then been observed in Malagasy. But now this seems to be corroborated by the discovery of the two forms in Malagasy. How far this is valid only for Malagasy, or for old Austronesian in general, is an open question.

Starosta, Pawley and Reid have this objection:

> Dahl cites Amis IF sa- as one justification for the initial *S, but it turns out that Amis sa- is not a regular IF marker in Amis. Instead, Amis sa- derives instrumental nominalisations which only rarely occur in a construction which could be analysed as having an Instrumental subject.

It is somewhat puzzling to find this as an objection in a paper which claims that the original function of this and other focus affixes was nominalisation (Starosta et al. 1982a:165; 1982b:131). The authors rely on Teresa Chen 1982: 117:

> Although the verbs in these examples ... have instrumental subjects, it would be somewhat misleading to label them "Instrument Focus" in the Philippine sense because they do not form a paradigm, and are not marked by any consistent IF affix.

Ferrell, however, gives a paradigm with sa- as morpheme of IF, following Ogawa and Asai 1935:403, although he also gives examples of the same kind as Chen (Ferrell 1972:122-123). It is possible that her investigations and those of Ogawa and Asai have been made in different Amis dialects. But at any rate Amis sa- does not seem to indicate a moving object, and is therefore not exactly parallel to Mlg a-.

In Saaroa, however, focus forms with the prefix saa- (which Tsuchida calls "special focus") may have as subject an instrument/means or an object, e.g. speaking in a fairytale of a bamboo on which a girl was climbing, it is said:

(12) Saa- ləvə-a ami muu-capi na ałaina isa
IF *go by means of is-said* AF *drop to woman her*
She came down by means of (it) to her mother, it is said.

(The function of the final -a, present in some forms but not in all, is not clear.)

(13) Saa-ɬamar-a cu ami ka tapuɬacəŋə
IF *burn already is-said the monkey*
(It) was burnt by the monkey, it is said. (Tsuchida 1976:75-77)

In (12) the means for climbing is in focus, in (13) the object burnt. This resembles the double use of a- in Malagasy, but the object is not clearly moving.

5.3 When Malagasy verbs in AF have two objects, the direct object is often moving, and the receiver of it is the indirect object. Both these may be focused, the direct one with the a- form, the indirect one with the -ana form, e.g. Mer

(14) manólotra fanomézana an- drai- ny ʃzy
hand over gift acc. *father his he*
He presents his father with a gift.

The same may be said in the following ways:

(14a) tolór-a-ny fanomézana (ny) rai-ny
His father is given a gift by him. Or

(14b) a-tólo-ny an-drai-ny ny fanomézana
The gift is given to his father by him.

Manólotra is AF, tolór-a-ny is RF tolór-ana + third person suffix, a-tólo-ny is IF a-tólotra + third person suffix, from the wordbase tólotra.

5.4 Because the IF forms begin with the vowel a, past tense has only n and future tense only h-, like these tenses in forms with -ina and -ana when the wordbase has initial vowel.

6. IMPERATIVE

6.1 The only mood different from the forms treated above is the imperative. The AF imperative is formed with the suffix -a to AF present tense, e.g. mi-sótro *to drink*, mi-sotró-a *drink!* Because a syllable is added, the accent is moved to the new penult. If the wordbase has final -ka or -tra, the suffixation of the imperative morpheme follows the same diachronic rules as that of -ina and -ana (cf. 1.2.1-1.2.2), e.g. mi-pétraka *to sit*, mi-petráh-a *sit down!*, manóratra *to write*, manorát-a *write!* When a wordbase with antepenultimate accent has final -na, the same rule leads to the following consequence: the imperative morpheme replaces the secondary final a, and this displaces the stress to the penult, e.g. manátona *to approach*, manaton + the imperative suffix -a = manatón-a *draw near!* The only audible and visible difference between AF present tense and imperative is then the place of the accent.

If a wordbase accentuated on the penult has a in final position, the -a of the suffix contracts with the preceding a, e.g. mi-ála *to go away*, *mi-alá-a > mialá *get out!* Here too the shift of the accent, now to the ultima, is the only difference between the two forms.

6.2 All the non-active focuses have the same imperative suffixes. In Merina the regular suffix is -o, but when there is an o in the wordbase, -y [i] is used instead of -o by a rule of euphony. Before the suffixation of -o or -y the suffixes -ina and -ana are deleted. OF and RF have thus the same form in imperative, e.g. tapáh-ina *to be cut*, tapáh-o ny tady *cut the rope!*, sorát-ana *to be written*, sorát-y ny taratásy *write the letter!*, a-tólotra *to be handed over*, a-tolór-y azy ny vóla *give him the money!* The "thing" in focus is subject of the verb in the imperative too, and must be in a definite form, here as in non-imperative clauses. The subject is not always pronounced, but nevertheless implicit, because it is "old information". If the object of the action is indefinite, active imperative must be used, e.g. mambóly váry *to plant rice*, mambolé váry *plant rice!* We see that if the wordbase had an original final e, Merina has crasis between this and the -a, as has been observed with -ana, while Sakalava has mambolé-a (cf. 4.2 above).

In Sakalava the suffix of the non-active focuses is always -o, e.g. sorát-o zao *write this!* When the wordbase has final o, the two o's are contracted, e.g. vonó-e *to be killed*, vonó *kill!* from vonó-o.

6.3 The suffix -i is the morpheme of RF imperative in Atayal too (Egerod 1965: 269), and also imperative in Sedek (Asai 1953:56). It is used in RF or locative forms in several Philippine languages. Mlg -o may be cognate to Atayal -au, which is OF subjunctive morpheme in this language (Egerod 1965:269), and has similar use in other Formosan languages, e.g. Paiwan.

7. VERB OR NOUN?

7.1 Since Dempwolff, it has been discussed whether the non-active focus forms are verbal or nominal. Dempwolff always spoke of *das nominale Denken der Austronesier* 'the nominal thinking of the Austronesians'. This is endorsed by Erin Asai (1936:37; 1953:62-63) and Cecilio Lopez (1941) who were among his students. The reason given by Dempwolff was that these forms construct the actor in the same way as the owner to his possession, and should accordingly be considered as nouns.

With my background in Malagasy I could not accept this. I had the feeling that the forms of non-actor focus were as verbal as AF, expressing actions and states to the same degree. But feelings are not scientific arguments, and for a foreigner semantics is too often influenced by translation to his mother-tongue.

In Indo-European languages we have verbal nouns expressing actions, but syntactically their function is nominal, even when they, as infinitive, have no nominal inflection. The decisive criteria must therefore be found in syntax. Is it possible to find such criteria in Malagasy syntax?

7.2.1 In a simple clause where no member is emphasised more than others, the word order in Malagasy is predicate (new information)-subject (old information). Both predicate and subject may be either noun or verb, e.g.

(15) mpam-bóly ny mp-ónina éto
farmer the inhabitant here
The inhabitants (P) *here are farmers* (S).

(16) mam-bóly váry ny mp-ónina
cultivate rice the inhabitant
The inhabitants (S) *cultivate rice* (P).

(17) mpam-bóly daholo ny m-ónina éto
farmer all the AF *live here*
They who live (S) *here are all farmers* (P).

(18) mam-bóly váry ny m-ónina éto
AF *cultivate rice the* AF *live here*
They who live (S) *here cultivate* (P) *rice.*

In clauses of this type the subject, whether noun or verb, must be in a definite form in Merina, whether preceded by the article ny, or preceded and followed by a demonstrative pronoun. (Proper nouns and pronouns are definite by nature and do not require these determiners.) In the examples above the AF verb monina is nominalised by the article, and functions like a participle, but without any formal change of the finite verb.

(19) mamonjy aina ny man-dósitra
AF *save life the* AF *flee, run away*

This may be understood in two ways, whether: *to flee* (S) *saves* (P) *life*, or: *they who flee save their lives.* The nominalised verb may thus function like an infinitive or like a participle in Indo-European. In these clauses the predicate, noun or verb, cannot be in a definite form with ny.

In the examples above I have used AF because this form bears no morphological resemblance to nouns. It takes objects and complements in the same way as active verbs generally do. It should therefore be possible to agree on their

verbal nature. As subject with ny we must consider them as nominalised verbs, deverbatives, but their form is exactly the same as the finite verb.

7.2.2 In the descriptions above of the three non-active focuses there are several examples of these forms used as predicate (no.1, 2, 3, 4b, 5, 6, 7b, 8b, 9, 11, 14a, 14b). But they may also be used as subject, e.g.

(20) tsy ázo- ko ny laza-in-ao
not understood I the say OF *you*
I do not understand what you say. (lit. *the said* (S) *by you (is) not understood* (P) *by me.*)

(21) mba vakí-o ny no- sorát-an-ao
please read imper. *the* past *write* RF *you*
Please read (P) *what you wrote* (S).

(22) rátsy tokóa ny a-fíndra ho amin' ío lásy ío
bad very the IF *move to that camp that*
To be moved (S) *to that camp is very bad* (P).

7.2.3 When the predicate is strongly emphasised, to the virtual exclusion of all else, the clause has a special structure. The subject (old information) is in Merina preceded by no (in other dialects by ro), particles resembling the case markers in Formosan and Philippine languages. It is very often natural to translate it in our languages as *it is X that/who is/does Y*. In Malagasy X is predicate and Y is subject. After no the article ny cannot be used (and only rarely after ro). Both predicate and subject may be noun or verb, and a noun as predicate may be in an indefinite or definite form, e.g.

(23) mpampiánatra no tómpo-n' ny tráno
teacher owner gen. *the house*
It is a teacher (P) *who is (the) owner* (S) *of the house.*

(24) ny mpampiánatra no tómpon'ny tráno
It is the teacher (P) *who is (the) owner of the house.*

(25) (ny) vazáha no mampiánatra tény frantsay
(the) stranger AF *teach word French*
It is a/the stranger (P) *who teaches* (S) *French.*

But if a verb is predicate, it is hardly ever possible to use it in an indefinite form. It must generally be preceded by the article ny. Compare the following examples:

(26) (ny) mpampiánatra tény frantsay no fántatr-o
(the) teacher word French known I
It is a/the teacher (P) *of French whom I know* (S).

As mpampianatra is a noun, it may be in indefinite or definite form. But here it is possible to use the verb mampianatra *teach* instead of mpampianatra *teacher*, and then it must be preceded by the article:

(26a) ny mampiánatra tény frantsay no fántatr-o
the AF *teach word French known I*
lit. *It is the teaching* (P) *French whom I know* (S).

Here the verb expresses the actor. But to say:

(26b) **mampianatra teny frantsay no fantatro
It is a teaching French whom I know.

would not be grammatical. To have an indefinite predicate we must introduce a noun to which mampianatra is an attribute:

(26c) ólona mampianatra teny frantsay no fantatr-o
person teach word French known I
It is somebody (lit. *a person*) *teaching French whom I know.*

If, however, the verb expresses the action and not an actor (cf. example (19)), it may be without article, e.g.

(27) (ny) mampianatra teny frantsay no raharáha-ny
It is to teach (P) *French that is his/her/their occupation* (S).

With or without an article the meaning is the same.

With verbs in the non-actor focuses the syntax is the same, e.g.

(28) ny n- irí- ko no éfa ázo-ko
the past *desire* OF *I finished got I*
It is what I desired (P) *that I have got* (S).

(29) ny éfa no- sorát-ana no ho- vakína
the perfective past *write* RF fut. *read* OF
It is what has been written (P) *that shall be read* (S).

(30) ny h- a- éli- ko no lazá-i-ko amin-ao
the fut. IF *diffuse I say* OF *I to you*
It is what I shall diffuse (P) *that I tell* (S) *you.*

In these cases the verb must be preceded by the article, because it is used like a participle expressing, e.g., the object of the action. But the non-actor focuses too may express the action itself, and then ny is facultative, e.g.

(31) (ny) arí- ana no antónona ázy
(the) throw away RF *suitable it*
It is to be thrown away (P) *that is suitable* (S) *for it.*

We may say that the most nominal use of the verb expresses the action itself (like an infinitive or a gerund) and is treated syntactically as a noun. Where it expresses actor or object (like a participle), it retains more of its verbal character and is treated unlike a noun. Actor focus and non-actor focuses are treated in the same way, as verbs, not as nouns.

7.2.4 Malagasy has yet another construction which clearly shows that in this language, at any rate, Dempwolff's argument is not valid as proof of nominal character. Non-active imperative forms may have the short-forms of the second person pronoun suffixed to them, like nouns and the ordinary non-active focus forms. The appeal is then emphasised. And imperative is incontestably a verbal, not a nominal form. Compare the following examples:

(32) trano-n- ao ity
house poss. *you*(sg.) *this*
this is your house

(33) tafó- an- ao ny trano
thatch RF *you the house*
you are thatching the house

(34) hevér-o- n- ao ny toétra- ny
think imper.poss. *you the situation his*
Do think (P) *of his/her/their situation* (S).

(35) Fidi- o- n- areo anío izay ho- tompo-in-areo
choose imper.pos. *you*(pl.) *today who* fut. *serve* OF *you*(pl.)
Do choose (P) *today whom you*(pl.) *will serve* (S).

7.2.5 Morphologically noun and verb have owner and actor constructed in the same way, but syntactically nominal and verbal forms have different constructions. The limit between the two categories is, however, not the same as in Indo-European languages. Used about the action (like an infinitive) the verb has nominal character, but used like a participle it is verbal. The participle is an adjectival form, and in Malagasy the adjective belongs to the verbal category. It is often formed with the verbal prefix ma- and has the same tense inflection as the verb, past na-, future ha-. It has also an imperative form with -a like AF, e.g. ma-dio *clean*, na-dio *was clean*, ha-dio *will be clean*, ma-diov-a *be clean!*

7.2.6 In Malagasy it thus seems clear that the non-active focus forms are verbal. Only a syntactic examination of other languages can show whether this is the case in these languages too. The criteria have to be chosen according to the syntactic rules of each language. A comparison of the results may inform us about the character of these forms in modern languages.

7.3.1 What are we able to say today about the situation in Proto-Austronesian? Starosta, Pawley and Reid argue that:

> *-ən, *ni-/-in-, *-ana, *-iSi-, and possibly *mu-/-um- were all noun-deriving affixes in PAN, as they still are to a large extent in the modern languages outside the Philippine area, and that they have in fact retained this function to a previously unrecognised extent even within the Philippine language group. We argue further that Austronesian nominalisations in *-ən, *ni-/-in-, *-ana, *iSi- and possibly *mu-/-um- did not develop from original passive constructions, as concluded by Dahl (1973), Wolff (1979), and Pawley and Reid (1979), but rather that the nominalising function was the original one, and that the passive and verbal focus uses of these affixes in Philippine languages are a secondary development. That is, verbal focus in Proto-Austronesian was at most an incipient mechanism that was later elaborated and developed by the languages of Borneo and the Celebes.
>
> (Starosta et al. 1982a:148)

We shall see that development from nominal into verbal forms has probably taken place in Malagasy (see 8.3 below), and such changes are certainly possible.

7.3.2 However, if we consider as PAN only the nominal forms with the affixes of the focus system, the focus system itself must have developed later. In my opinion this is not possible. The focus system is found both in some Formosan and in some western languages from the Philippines to Madagascar with similar

forms and in similar constructions. But the languages where it is found belong to different primary subgroups of Austronesian. The Formosan languages have so many archaic features which they do not have in common with languages outside Formosa, that they must represent the first offshoots from the PAN centre (Dahl 1973:124-125; 1976:125; 1981:153; Blust 1980:13). Till now I have not been able to find innovations common to all Formosan languages. For the time being they must therefore be considered as belonging to several first-order subgroups of Austronesian.

But all the languages outside Formosa have innovations in common, e.g. PAN $*S_1$, S_2, H_1, H_2 into PMP *h, and after this PAN *t' generally into s (Dahl 1981: 45-62). In the Formosan languages which have had PAN *t' > s, this s has merged with s < PAN $*S_1$ (1981:84). If s from both these PAN phonemes had existed simultaneously in the proto-language of the languages outside Formosa, we should have expected the same merger in at least some of these languages, but this is found nowhere outside Formosa. The change of PAN *t' into s outside Formosa must therefore have taken place after the change PAN $*S_1$ > h (Dahl 1981:87).

The nasalisation of the first consonant of the wordbase, which has lead to nasal accretion in Oceanic and nasal substitution in Western Austronesian, is also confined to the non-Formosan languages. We must therefore consider all languages outside Formosa as one primary subgroup, which Blust has called Malayo-Polynesian (see e.g. Blust 1980:13).

7.3.3 The four-focus grammatical system exists both in Formosan and MP languages, that is to say in more primary subgroups. If this system did not belong to PAN but developed later from nominal PAN forms, parallel development must have taken place in several subgroups after their separation. However, not only are the morphemes identical, but also the syntactical use of the forms. The choice of focus form permits placement as subject words with different relations to the action or state expressed in the clause. A parallel development of this sophisticated system in different subgroups of AN from nominal forms with the same affixes does not seem possible to me. There are too many similarities. For instance, in Atayal and Malagasy the fundamental features of the four-focus system are virtually the same in spite of the long separate development of the languages. The only significant difference is that the more differentiated modal categories in Atayal are reduced to two in Malagasy.

To me the possibility of parallel development from verbal into nominal forms seems much greater. The uses of the non-verbal forms are not so similar in the different languages. In Malagasy these forms are generally not nouns but adjectives, and thus nearer the quality of the verb. The Malagasy focus forms are not only used as subject and predicate, but also as qualifiers of nouns, like adjectives, e.g. závatra oména *a thing given*, in constructions like závatra mavésatra *a heavy thing*. From such constructions to the use of the same affixes to create adjectives the way is short, e.g. ráno lomór-ina *water overgrown with moss* from lómotra *moss*, tány vató-ana *stony earth* from váto *stone*. (Neither lomotra nor vato have verbal forms.) If the noun is omitted in such constructions, the adjective is nominalised. The next step in the evolution may then be a real noun.

We have seen (7.2.5 above) that in an AN language the difference between noun and verb is not the same and not so sharp as in IE languages. For that reason, the possibility that the same affixes as those forming focus may have been used to produce nominal forms in PAN itself cannot be excluded (cf. Saaroa saa- in 5.2.3 above).

7.3.4 If the focus system belonged to PAN, some Formosan and many MP languages have lost it, among others Malagasy's nearest relatives, the Barito languages. Here about 1600 years of separation has been sufficient to produce the difference. PAN was probably spoken around 5000 B.C. (Blust 1980:13), and a loss of fundamental grammatical features during these 7000 years is not extraordinary. Modern European languages have lost the fundamental case structure of Proto-Indo-European in a much shorter time, replacing it with a set of prepositions formed from old material. In languages which have lost the focus system, there may be new forms replacing categories in the focus system, here also using old material. It may be worth examining modern grammatical systems with this in mind.

8. CIRCUMSTANTIAL FOCUS

8.1 Malagasy has also a fifth focus which is formed by a circumfix where the suffix is always Mer -ana, Sak -a, and the prefix is any AF prefix deprived of its initial m-. The form has thus initial vowel and therefore n- and h- in past and future tenses like IF. In most dialects it forms its imperative according to the same rules as the other non-active focuses. But in Tesaka the imperative of this focus always has the suffix -y (Deschamps 1938:20). We have already seen that -i forms the imperative of RF in Atayal (6.3 above), and in other languages it is a locative suffix. Because CF has the suffix -ana, it has some resemblance to RF. This may be the reason why -y is the imperative suffix here, and this may be the origin of the alternative -y in Merina too.

This focus was called relative voice by the old grammarians (Cousins 1894: 48) because it has in focus any relation to the action. However, since this form has no similarity to the accustomed use of relative in grammars (relative pronoun, relative clause), I prefer another term: circumstantial focus (CF).

8.2 Any circumstance having a relation to the action or state expressed by the verb may be focused by this form: place, time, cause, intention, reason, means, instrument, beneficient, e.g.

(32) i-petráh-a-ny ny séza (petraka *to sit*)
sit CF *he the chair*
The chair (S) *is where he is sitting* (P).

(33) i-angón-ana ny alahády (ángona *to assemble, go to church*)
assemble CF *the Sunday*
Sunday (S) *is when going to church* (P).

When CF is used, there is often so great an emphasis on the circumstance that the construction with no is preferred:

(34) táhotra no n- an-dosír-a-ny (lósitra *flight*)
fear past *flee* CF *he*
It was for fear (P) *that he fled* (S).

(35) mba h- amotsí-ana ny tráno no ilá- ko sokay
conj. fut. *white* CF *the house* *want I lime*
It is in order to whitewash (P) *the house that I want* (S) *lime.*

(The wordbase of hamotsiana is fótsy *white*, AF mamótsy, and of ilako íla, AF m-íla *to want*. Because the AF prefix is only m-, which must be deleted, CF of this verb has no prefix.)

(36) sokay no h-amotsí-a-ko ny trano
It is with lime (P) *that I shall whitewash* (S) *the house.*

(37) ny ántsi- n- ao no anapáh- o ny tády
the knife gen. *you cut* CF imp. *the rope*
Cut (S) *the rope with your knife* (P).

(38) ny ray aman- drény no h- anolór- an- ao ny vóla
the father and mother fut. *present* CF *you the money*
It is to the parents (P) *that you shall give* (S) *the money.*

The CF may also express an action on only a part of the subject. Compare the following clauses:

(39) vono- y ny akóho- ko
kill OF imper. *the chicken my*
kill my chickens (all of them)

(40) amono-y ny akóho-ko
kill some of my chickens

(41) amono-y róa ny akóho-ko
kill two of my chickens

Ny akohoko is the subject of all the three clauses, róa in the last one is object of amonoy. The command in the last one may also be expressed in AF, but in a less elegant manner:

(41a) mamonó- a róa ámin' ny akóho- ko
AF *kill* imp. *two among the chicken my*

The fact that the part in AF is expressed with a complement (amin'ny a-), and accordingly as a circumstance in relation to the verb, explains the use of CF with the same meaning.

Preceded by the article the CF may also mean the action itself nominalised, e.g. ny i-petráh-ana *the sitting, the action to sit.*

The CF form makes the language very flexible in that it allows any part of a statement to be emphasised. This is necessary because more than ordinary stress cannot be used to emphasise single words. Focus thus offers a suitable means of achieving emphasis.

8.3 I have found no clear parallel to CF in any other AN language. It therefore seems to be a Malagasy innovation. Looking for material from which it has been created, I have found some forms in Ma'anyan. This language has some abstract nouns formed with the circumfix pa- with nasal substitution or accretion + -an or with pi- + -an. The meanings of these forms are the following: the action itself, the place, time, instrument (Sundermann 1913:219-221).

Ma'anyan p has become Mlg f, and in Malagasy we have the same forms with f-: fa- with nasal substitution or accretion + Mer -ana, Sak -a, and fi- + -ana/a. The meanings of the Mlg forms are the same as in Ma'anyan plus most of the meanings of the CF. But the Mlg forms with f- and the Mny forms with p- add to these meanings the notion of habitual. Compare the following:

(42) ny fanaov-ana ázy (from tao *to do, make*)
the doing it
The habitual, general way of doing it (always).

(42a) ny anaov-ana ázy
the way of doing it (in the actual situation).

There are also forms without -ana which have the same habitual meaning, e.g. f-ómba *custom, habit* from ómba *to accompany*.

It is therefore possible that the old forms with p/f + -an/ana and habitual meaning have eliminated the f-, and thereby removed the connotation of habitude. What is left is the form of CF, and to this still other meanings have been added. For the time being I consider this to be the most likely hypothesis.

It is worth noting that the nominal form with f- has an object, just like the verbal CF form without f-. This shows that in Malagasy the difference between noun and verb is not so sharp as in Indo-European. It would be interesting to know if this is true of other AN languages.

9. CONCLUSIONS

Malagasy has the four-focus grammatical system which is also found in Formosan, Philippine and Minahasan languages, and with affixes that are present in other focus languages too. It is found here that the PAN forms of these affixes are AF *-um-, OF *-ən and -in-, RF *-an and IF *S_1i- (perhaps also *Sa-). Malagasy shows regular reflexes of all these. The imperative suffixes present in Malagasy are AF -a < PAN *-a, in the non-active focuses -i < PAN *-i and -o < PAN *-au. Instead of PAN *-um- Malagasy mostly uses reflexes of *ma- + nasal accretion or substitution, or of *maɣ-.

Since Dempwolff the question of whether the non-active focuses are verbal or nominal has been a moot point. Malagasy syntax shows that in this language they are verbs - in spite of the construction of the actor being in the same form as the owner of the noun expressing his possession. The non-active imperatives may suffix the second person short form, like nouns. But imperatives are incontestably verbal forms. This shows that such construction is no proof of nominal character. To settle this question for the focus languages in general a syntactical examination of the function of focus in these languages is needed.

Today the focus system is found in several first-order subgroups of Austronesian (see 7.3.2 above). The similarities are so great that a parallel development of the system in these languages must be excluded. Focus must therefore have belonged to the PAN grammatical system. Malagasy has, however, developed a fifth focus which must be a local innovation.

The languages without the focus system today must have lost it. Instead there are sometimes found new forms replacing categories in this system, formed at least partly with old material. A further study of this in Austronesian languages is needed.

NOTES

1. The following abbreviations are used: AN = Austronesian, Isb = Isbukun, Mer = Merina, Mlg = Malagasy, Mny = Ma'anyan, MP = Malayo-Polynesian, P = predicate, PAN = Proto-Austronesian, PMlg = Proto-Malagasy, PMP = Proto-Malayo-Polynesian, S = subject, Sak = Sakalava, Tkb = Takbanuad, Ttd = Takituʔduh.

2. In Malagasy orthography o is the symbol for the vowel [u], and y is written for final [i]. I mostly take my examples from Merina, which is the base of the official literary Malagasy, but also from Sakalava, and only occasionally from other dialects, when they give us information about the historical development of the language.

3. There are two possible proto-forms for this wordbase: PAN *laɣiu and *laiat', both meaning *to run*, and both with irregular development of the last vowel. The two have probably merged, because in imperative we have both l-om-ai-a and l-om-ais-a, cf. 6.1.

4. That there has been a prefixed form with mu- is corroborated by the causative form of these verbs. The ordinary causative prefixes in Malagasy are mampa- + nasal accretion or substitution, or mampi-, corresponding with the AF prefixes ma- + nasal accretion or substitution or with mi-. But in Sakalava the causative prefix of the verbs with -om- is mampo-, e.g. mampolay *to cause to run*. See also mu- in Formosan languages (Dahl 1973/76:119).

5. For the development of final nasals, contraction of vowels and accent in Malagasy, see Dahl 1951:62-65 and 84-91.

6. omé < PAN *bəɣai with fossilised -um-: *b-um-əɣai > *wumée > omé. Only the loss of w before u is irregular.

7. Ferrell has misunderstood two forms in Egerod's paradigm. In private correspondence Egerod has informed me that perfective OF should be .q-n-alup and RF q-nalup-an. In perfective IF the form Egerod has written (i)nsqalup is so rare in his material that he does not consider it as certain. Ogawa and Asai do not have this form in their paradigm (1935:30).

REFERENCES

ASAI, Erin

1936 *A study of the Yami language, an Indonesian language spoken on Botel Tobago Island*. Leiden: J. Ginsberg.

1953 *The Sedik language of Formosa*. Kanazawa: Cercle Linguistique de Kanazawa.

BLUST, Robert A.

1980 Austronesian etymologies. *Oceanic Linguistics* 19:1-189.

CHEN, Teresa M.

1985 *Verbal constructions and verbal classification in Nataoran-Amis*. (Ph.D. dissertation, University of Hawaii, 1982.) *PL*, C-85.

COUSINS, W.E.

1894 *A concise introduction to the study of the Malagasy language as spoken in Imerina*. Antananarivo: The Press of the L.M.S.

DAHL, Otto Chr.

1951 *Malgache et maanjan. Une comparaison linguistique*. Oslo: Egede-Instituttet.

1954 Le substrat bantou en malgache. *Norsk Tidsskrift for Sprogvidenskap* 17:325-362.

1971 La forme locative du nom malgache. *Norsk Tidsskrift for Sprogvidenskap* 24:154-162.

1973 *Proto-Austronesian*. Scandinavian Institute of Asian Studies, Monograph Series No.15. Lund: Studentlitteratur. 2nd revised edn, 1976.

1977 La subdivision de la famille Barito et la place du malgache. *Acta Orientalia* 38:77-134.

1978 The fourth focus. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, 383-393. *PL*, C-61.

1981 *Early phonetic and phonemic changes in Austronesian*. Oslo: The Institute for Comparative Research in Human Culture B 63. Universitetsforlaget.

1983 *Sorabe révélant l'évolution du dialecte antemoro*. Antananarivo: T.P.F.L.M.

DESCHAMPS, Hubert

1936 *Le dialecte Antaisaka*. Tananarive: Pitot de la Beaujardière.

EGEROD, Søren

1965 Verb inflection in Atayal. *Lingua* 15:251-282.

FERRELL, Raleigh

1972 Verb systems in Formosan languages. In Jaqueline M.C. Thomas and Lucien Bernot, eds *Langues et techniques, nature et société*, vol.1, 121-128. Paris: Klincksieck.

LOPEZ, Cecilio

1941 *A manual of the Philippine national language*. 3rd edn. Manila: Bureau of Printing.

OGAWA, Naoyoshi and Erin ASAI

1935 *The myths and traditions of the Formosan native tribes*. Taihoku: Taihoku Imperial University.

SNEDDON, J.N.

1978 *Proto-Minahasan: phonology, morphology and wordlist*. *PL*, B-54.

STAROSTA, Stanley, Andrew K. PAWLEY and Lawrence A. REID

1982a The evolution of focus in Austronesian. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 145-170. *PL*, C-75.

1982b The evolution of focus in Austronesian. MS. An expanded version of 1982a.

SUNDERMANN, Hermann

1913 Der Dialekt der "Olon Maanjan" (Dajak) in Süd-ost-Borneo. *Bijdragen tot de taal-, land- en volkenkunde van Nederlandsch-Indië* 66:203-236.

TSUCHIDA, Shigeru

1976 *Reconstruction of Proto-Tsouic phonology*. Study of Languages and Cultures of Asia and Africa Monograph Series No.5. Tokyo.

TUNG, T'ung-ho

1964 *A descriptive study of the Tsou language, Formosa*. Institute of History and Philology special publication No.48. Taipei: Academia Sinica.

ZORC, David Paul

1977 *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. *PL*, C-44.

Dahl, O.C. "Focus in Malagasy and Proto-Austronesian". In Geraghty, P., Carrington, L. and Wurm, S.A. editors, *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*.
C-93:21-42. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.21*

ON THE PRAGMATICS OF FOCUS

Paz Buenaventura Naylor

1. INTRODUCTION

In English, we can talk about the Tigers winning the baseball championship in at least four different ways. We can say:

1. The Tigers won the baseball championship.
2. The baseball championship was won by the Tigers.
3. It was the Tigers who won the baseball championship.
4. It was the baseball championship that the Tigers won.

These sentences all refer to the same event - the same extralingual reality. What, then, is the difference? The English speaker can readily sense that the difference lies in the context of situation in which each would be appropriate. When or in what situation do we say it one way and when another? It is easy to see that there are contextual constraints on the latter three and which one is used depends on what was said before or *what* the interlocutors were talking about. On the other hand, sentence 1 is more neutral. One could start a new topic of conversation with it. However, if one said "The baseball championship was won by the Tigers" it implies that the listener already knows that the baseball championship was being played and that the speaker was simply telling him *who* won. The third and fourth sentences are both contrastive. "It was the Tigers - not the Yankees (or some other team) - who won the baseball championship". Similarly, to say "It was the baseball championship that the Tigers won" implies that it was not the football championship - in case you did not know that the Tigers were a baseball team - that the Tigers won.

This is precisely the sort of thing I have been trying to figure out for Tagalog. In Tagalog, focus[1] selection works along similar principles, although the mechanics and the discourse dynamics are different. Similar forms function differently and similar functions are realised by different forms. Furthermore, within the rubric of 'passive form' in Tagalog, other noun arguments may be made the 'subject' or the 'in-focus NP' of the sentence, accompanied by the appropriate verb morphology.

Some linguists have said that, unlike English, the goal-focus sentence type that parallels the English passive (sentence 2 above) in its form, occurs more frequently than the actor-focus type (the 'active' like sentence 1 above). It has also been said that the goal-focus construction, although formally like the English passive, does not function in Tagalog in the same way that the passive does in English.

If these observations are accurate, as they seem to be, the obvious question is: Why is it so? The obvious answer appears to be: The choice of focus construction depends on *what* we are talking about. When we are talking about an

---

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 43-57. *Pacific Linguistics*, C-93, 1986.

event, the Tagalog system provides multiple options for grammaticalising the *focus of orientation of talk: not* just the actor or goal (agent or patient) but also the location, instrument, or beneficiary may be the in-focus NP. This means, then, that the speakers have a wider choice as to which facet of an event or situation talk must be oriented to. Furthermore, when the goal (patient or direct object) is 'definite', specific, or referential (recoverable in context), it has to be encoded as the surface subject of a goal-focus (formal passive) sentence.[2]

In the past ten years or so, especially since the first international Austronesian conference (Hawaii, 1974) when Austronesian scholars from different traditions came together perhaps for the first time, parallel insights and the outlines of some form of underlying unity in the pragmatic systems of Austronesian languages have begun to emerge. The sense of what the pragmatics consisted of, however, remained largely implicit in the descriptions of the syntactic structures. It became clear nonetheless that *focus* (in the Philippine linguistics sense), or some form of marking in the verb that the associated NP was 'in focus', ultimately motivated contrasts of *voice, aspect,* or *transitivity*. The fact that formally passive sentences in certain Austronesian languages are not necessarily 'real' passives (Milner 1974 and Tchekhoff 1974) and that they function in discourse differently from the English passive (McCune 1979) has been convincingly argued in the literature.

Several studies, based wholly or in part on Philippine language data, have also brought out the relationship between *focus* selection and discourse organisation (among them, Pike 1962, 1963; Naylor 1973, 1974, 1975; Hopper and Thompson 1980). They have brought out the important fact that *focus selection is constrained by the context*[3] *and by the function of sentence types in the flow of discourse*.

The difference in discourse function of the Tagalog goal-focus (the formal passive) construction is reflected in the fact that such a 'passive' construction occurs with much higher frequency in Tagalog discourse than the English passive does in English discourse. It has even been argued that it is the passive, not the active, construction that is basic in Tagalog (Ceña 1977).

Yet, to date, a relative frequency study of the occurrence in discourse of the goal-focus (GF) construction has yet to appear. The need for this sort of 'concrete' evidence has long been felt, however (Kess 1979). Such a study should confirm or deny the oft-repeated statement (as yet unsupported) that the GF construction is of equal or greater frequency than the actor-focus (AF) or 'active' construction in Tagalog.

Obviously, counting just for the sake of counting may be of little or no value. But not only must we count occurrence of focus constructions but we must also *account for their pragmatic motivation*. Thus, as a tool for the analysis of the dynamics of discourse, a relative frequency study provides an irrefutable and revealing record of usage on which we can base our analyses and interpretations with a good measure of confidence. Furthermore, as the discussion below will show, relative frequency data yield a number of extremely valuable insights into discourse organisation that may otherwise not become apparent. Certain such insights are of even greater importance than simply providing conclusive evidence for certain intuitions or impressions (e.g., as pointed out above, the higher incidence of GF constructions in Tagalog). At the same time, relative frequency data will, hopefully, enable us to describe these insights in more explicit terms.

What the present study proposes to do, therefore, is to seek additional insights into the pragmatics of *focus* on the basis of a relative frequency count

of AF and GF sentences in different types of discourse and in different types of text that typically occur at certain points in the discourse.

In Naylor 1973, 1975, a pilot study of relative frequency of occurrence of the various focus construction types, as they occurred in a variety of selected texts, provided the data for the study of *topic, focus,* and *emphasis* in the Tagalog verbal clause. The frequency figures from that earlier study were not reported at that time. However, they have now been combined with the frequency figures from the present study. (Tables 1 and 2 show the combined frequency figures except when otherwise noted.)

The purpose of the earlier pilot study was limited to the determination of the overall frequency of occurrence of the various focus constructions and the relative frequency of occurrence of these same focus types at *points of introduction of new subject matter*, taking into account certain contexts of situation in which certain subtypes occurred. The present study on the other hand is an expansion of the earlier one in terms of a much enlarged corpus of material and of more detail in the description of the observed occurrences of AF and GF construction types. It is also an expansion in the sense that we have developed greater capacity for deeper insights into the dynamics of focus as well as the dynamics of discourse in the intervening years. There is a marked contrast in the state of the art of discourse analysis between 1971 and 1984. We can now avail ourselves of a large body of literature - a veritable groundswell that began in the late 1970s - that deals with virtually all aspects of text study and discourse analysis.

It is interesting to note at this point that the patterns that emerged as a result of the earlier pilot study were essentially paralleled by the patterns that emerged from the present study. One wonders if simply *more* material would yield any more valuable information than we already have. Perhaps, rather than simply increasing the data base, it might be more fruitful to analyse in greater detail the occurrence of AF and GF clauses, *as well as nonverbal* clauses, in relation to their position in the discourse and then, position by position, across the different genres, styles, and registers. We should then be able to define paradigmatic and syntagmatic relationships among the major sentence types, relative to positions in discourse and type of discourse.[4]

2. METHOD

The material used for the present study included: stories, stories in comic format (verbal-pictorial mode with preponderance of dialogue), stories in play format, articles and essays, newspapers (unlike English newspapers, these consisted of only a few tabloid pages), regular columns in magazines or newspapers, letters, a television newscast, and taped conversations.

All forms that carry a focus marker were tabulated and categorised as to syntactic function, i.e. nominal, verbal, adjectival, adverbial. Whenever a focus-marked form functioned as a verbal predicate, its clause was tabulated as a verbal clause and further classified into AF or GF. Whenever a focus-marked form occurred as a predicate noun or adjective, its clause was tabulated as a non-verbal clause. (See Table 2.) Furthermore, the tabulations were distinguished as to position in the discourse: whether it was introductory or initial, developmental or medial, and for some texts, also if it was closure or final. (See Table 1.)

Operating on the hypothesis that 'intransitive' or AF clauses tend to code 'backgrounded' material and new information (therefore introductory material)

and that 'transitive' or GF clauses tend to code 'foregrounded', referential (old information) or developmental material, each clause that did not fit into this categorisation was *noted for features* that may help to explain the *motivation* for its use. In other words, we have assumed, a priori, on the basis of what we know about the syntax and semantics of AF versus GF sentence types, that AF is the unmarked choice for introductory material and GF for developmental material. Therefore, when marked choices occurred, it was also assumed that they were motivated by semantic, syntactic, pragmatic, or sociocultural constraints.

'Relative clauses' were not counted as verbal clauses since these are arguably neither verbal in function nor clausal in structure. The construction that most other linguists have analysed as 'relative clause' is, in Tagalog, actually a modifier phrase. It is identical in form and function to the Tagalog modification structure.[5]

A separate table (Table 3) identifies more categories of text within a particular discourse format. It was deemed necessary to do this for this particular story - a popularisation in comic format of a serious novel. The condensation of the novel required more narrative interludes to cover intervening events between the highlighted scenes presented in dialogue-pictorial form.

In the course of doing the present study, the high frequency of nonverbal sentences, earlier noticed but hitherto unstated, came forcibly to our attention.[6] For the present study, therefore, nonverbal-sentence frequency was also tabulated but only to show overall frequency of occurrence in the various genres and types of text, relative to the AF and GF sentences and their combined totals (see Tables 2 and 3 below). Since the present study is about the pragmatics of focus, and focus is a defining feature of verbal sentences in Tagalog, further detail on nonverbal-sentence occurrence did not seem relevant for the purposes of this paper.[7] The tabulation of nonverbal sentences was therefore done without regard for details of functional distribution within the various genres. On the other hand, since nonverbal sentences are part of the larger discourse picture, their frequency of occurrence relative to that of verbal sentences does provide the breadth of perspective that only such information can delineate. Such a perspective is necessary for a fuller and deeper understanding of the pragmatics of focus itself. In fact, it has now become clear that unless the functions of nonverbal sentence types are taken into account, we cannot have a fully adequate description of the pragmatics of focus (as realised by verbal sentences).

Finally, it should be pointed out that the frequency data from the earlier (1972) pilot study were combined with those of the present study. Since they were not perfectly parallel in regard to the categories distinguished and tabulated, appropriate notes regarding discrepancies are given in the tables.

3. DISCUSSION AND INTERPRETATION OF FREQUENCY DATA

3.1 Overall relative frequency

As we can see in Table 1, it is indeed the fact that in terms of total frequency of occurrence across various genres and types of text, and without taking into account the variations by genre and type of text, *GF sentences occur more frequently than AF sentences*. Note, however, that it is not that much higher in frequency of occurrence than the AF, as the total figures on the lower right-hand corner will show at a glance.

3.2 Introductory text

I have grouped together under the category of 'introductory text' sentences that occur as titles or headlines, as the initial sentence of the body of the discourse, introducers of a new topic of discourse or starters of a new episode.

It appears to be the consensus of opinion among several analysts of discourse organisation that sentences of the *actor-focus* type tend to introduce *new* topics of discourse and/or, present background information; they are seen as being on the lower end of the transitivity scale, thus tending to be *descriptive* rather than *narrative* in character (cf. Hopper and Thompson 1980).

The total figures in Table 1 for the relative frequency of occurrence of AF and GF sentences in introductory text confirm the above hypothesis. AF sentences occurred far more frequently. Yet if we examine the breakdown by genre, we can observe that the higher incidence of AF sentences is true of *only three genres*: stories (narratives), play format, and actual conversations. Furthermore, the ratio of AF to GF sentences in these discourse types was extremely high: almost 2 to 1 in narratives, and more than 4 to 1 in both the play format and the conversations. In the *other six genres*, however, *GF sentences showed higher incidence* over AF sentences: by a slight margin in the comic format and articles and essays, by a roughly 3 to 2 ratio in columns and the television newscast, and by 2 to 1 in newspapers and letters.

Already two important considerations emerge from this examination of the data. First, it is clear that *overall relative frequency figures* for AF and GF sentences do not reveal any more than the *simple ratio* of AF to GF sentences in discourse; they do *not* tell us about *how* focus functions in discourse. Secondly, this emphasises the crucial importance of *distinguishing types of discourse or genres* in the study of discourse pragmatics, in which the pragmatics of focus figures as one of the component 'networks' or systems. (At this point, we must bear in mind that the hypothesis stated above was formulated with specific reference to narrative text in Hopper and Thompson 1980.)

As we shall see in the discussion of the data with reference to developmental material, these considerations are just as important for the study of the pragmatics of focus in other types of text that occur at other points in the discourse, or that play a different role in the flow of discourse.

The taped conversations do confirm the hypothesis that AF sentences tend to introduce new topics of talk. From some of the things we know off-hand about the nature of conversation, it is easy to see why the ratio of AF sentences in introductory text should be so high. There is no narrator to mediate and the interlocutors have to continually provide background and transition themselves.[8]

The play format would parallel conversation and the narratives of course typify the basis on which Hopper and Thompson formulated their hypothesis. That these should show a high ratio of AF sentences in introductory material needs no further explanation, given the generally accepted function of AF sentences as coders of new information.

On the other hand, given the generally accepted notion that the GF sentence type 'refers' to something 'known' or 'given' or 'recoverable' somewhere in the context, that is 'referential' and therefore *does not convey new information*, it would be puzzling that GF sentences should occur in introductory material at all. Not only does it function to introduce new information, but it even does so with greater frequency than the predictable AF sentence type in a good number of genres. We are therefore challenged to account for *when* and *why* GF sentences occur in introductory text.

Table 1: relative frequency of actor-focus and goal-focus

	Stories	Comics Format	Play Format	Articles Essays	Columns	News-papers	Letters	TV News-cast	Convers-ations	TOTALS
Intro-ductory										
AF	224	77	210	61	11	60	2	26	201	872
GF	127	83	49	66	15	130	4	39	46	559
Develop-mental										
AF	217	438	290	189	67	53	24	17	80	1375
GF	349	471	478	237	123	185	22	31	259	2155
Closure										
AF	8*	1**	-	6***	7	-	3	-	-	25
GF	5*	2**	-	9***	3	-	6	-	-	25
TOTALS (does not include figures for 'Closure')										
AF	441	515	500	250	78	113	26	43	281	2247
GF	476	554	527	303	138	315	26	70	305	2714

*based on 35% of the material; **based on 68% of the material; ***based on 50% of the material

Table 2: Relative frequency of actor-focus, goal-focus, and nonverbal clauses in various genres

	Stories	Articles Essays	Single Topic	Columns	News-papers	Letters	TOTALS
AF	263	132	109	85	30	29	648
GF	309	178	111	141	73	32	844
Nonverbal	340	163	188	156	54	115	1016

	Comics Format	Play Format	TV News-cast	Convers-ation	TOTALS
AF	383	500	43	186	1112
GF	446	527	70	208	1251
Nonverbal	602	830	28	666	2126

First of all, we need to refine our concept of the *given-new* opposition. This has been sufficiently dealt with in the literature (e.g. Halliday 1968, 1973; Naylor 1973, 1974, 1975; and others). Yet it never seems to be superfluous to call attention to the fact that the given-new distinction is not necessarily realised by *discrete* forms or parts of sentences and that the concept is a *complex* of networks - multisystemic and multidimensional, neither simplex nor unidimensional. Thus, a sentence constituent may be *given* within one system of contrast and *new* within another system of contrast in the network. Without going into digressive detail at this point, suffice it to say that Naylor (op.cit.) has argued that in Tagalog, the pragmatic opposition lies in the *general-particular* rather than the *given-new* contrast. It follows, therefore, that: while GF sentences are undeniably referential, they may nonetheless convey *new* information, the *newness consisting of particularisation, contrastiveness, or the mere fact of being newly brought into the focus of attention* of the interlocutors.

These two concepts, referentiality and the general-particular contrast, as realised in the GF sentence type, provide options and linguistic resources that are fully exploited in journalistic text. (It should be noted again at this point that these were the discourse types in which the GF construction occurred with much greater frequency in introductory material than the AF sentence type.) In the examples that follow, the use of *marked* order reinforces the referentiality quotient, making it more pronounced and its immediacy greatly heightened.

As headings for news items, the newspapers in particular showed a great abundance of sentences of the following structure:

(1) Piskal, pinagdusa ni Marcos.
fiscal penalised
Fiscal (prosecuting attorney), penalised by Marcos.

(2) Custodio, 50 pa, pinawalan.
Custodio 50 yet freed
Custodio (proper noun), *50 others, freed.*

(3) Tsuper binaril ng pulis.
driver shot policeman
Driver shot by policeman.

There is a 'general rule' in Tagalog, by which the speaker is constrained to use the GF construction whenever the patient or 'direct object' or goal is 'definite'. Thus, the AF sentence,

Bumili si Pedro ng sapatos.
Pedro bought (a pair of) shoes.

contrasts with the GF sentence,

Binili ni Pedro ang sapatos.
Pedro bought the shoes.

The implication in the GF sentence is that Pedro bought *the* shoes - that we know about, that we had talked about earlier.

The GF sentence conveys the assumption of shared information, either through communication or through shared knowledge and shared experience in the shared cultural context. This shared information provides the basis for the 'definiteness'. As a discourse notion, 'definiteness' derives from the recoverability of the referent of the 'definite' NP in the context, linguistic or extra-linguistic, and not simply the old grammar-book definition of what the definite article indicates (cf. Naylor 1984).

We might point out here that, contrary to previous general impressions, recoverability of the referent is not necessarily in the preceding context; i.e. it is not necessarily anaphoric. It may be 'cataphoric'. In fact, *all discourse-initial sentences are necessarily 'cataphoric'* since they direct attention to what follows. Discourse-initial sentences also introduce what is to be developed - defined, circumscribed, or expanded in subsequent text. The referent may in fact by recoverable in *subsequent* rather than in preceding text (cf. Naylor 1985).

Given these observations about the GF construction and the pragmatic features associated with it on the one hand, and the thematic nature of discourse-initial and certain sentence-initial constituents on the other, it should now be easy to see what pragmatic motivations may account for the occurrence of GF sentences in introductory material in Tagalog discourse.

Let us now turn to some examples from the material used in the present study.

For example, in one of the taped conversations, the interlocutors were first cousins whose aunt left them an inheritance in her will. When the will was first introduced in the conversation, it was as the in-focus NP ('subject') of a GF sentence.

When someone is being introduced to someone else, the introduction is phrased as a GF sentence because it is 'known', by his or her presence, *who* is the subject of the introduction.

Yet another context in which the GF construction occurs discourse-initially is in the modern short story. Quite often, the story begins *in medias res* and we find a GF sentence opening the discourse. The author *presumes* that the reader *knows what* he or she is talking about and if the reader does not know at that initial point, the reader will know as the story unfolds - another example, and a common one, of cataphoric referentiality.

3.2.1 Performatives, quotatives, and adversatives

Performative and quotative sentences usually focus on *what* was said, thought, or promised, etc.; therefore, the verb of the Tagalog performative or quotative sentence tends to be in goal focus. (The promise or quotation is not coded as an *ang*-NP however.) This means that if the introductory sentence in the discourse is a performative or a quotative, it will usually be a GF construction.

There are a good number of verbs in Tagalog that are transitive in form but intransitive in meaning. They generally convey a state of affairs that may in one sense or another be considered adverse in its effect (cf. Dardjowidjojo 1979). For example:

(1) Baka kulangin tayo ng pambudbod.
might be short of we of topping
We might be wanting for (something to use) for topping (on the cake).

(2) Minalas si Juan.
bad-lucked Juan
Bad luck befell Juan.

These verbs are rarely if ever used in other than the goal focus, regardless of whether they are introducing a new topic of discourse or not.

3.2.2 The comic format

This genre shows a slightly higher frequency for GF sentences in introductory material. Comic stories, however, are a breed apart because there are pictures that complement or supplant words. In fact, I have observed in a full-length (52-page) comic story that precisely when the action gets intense, the picture says it all and there are no words that appear other than "Bang! Bang!" or "Oops!", etc. The narrator comes on rather frequently and this brings in narrative sequences to a format that would otherwise be like a play or dialogue simulating natural conversation. When it comes to introductory text, the comic

format shows an interesting divergence between narrative and conversational segments of the discourse in the matter of which focus construction occurs more frequently. The narrative introductory material shows a preponderance of AF constructions whereas the conversational introductory material shows a preponderance of GF constructions.

Because of its uniqueness, I have shown in Table 3 the breakdown of the figures given in Table 1 for this genre.

Table 3: Relative frequency of occurrence of actor-focus, goal-focus and nonverbal clauses in the comic version of *Noli me Tangere* (a famous novel written by José Rizal in Spanish, subsequently translated into English and Tagalog)

	AF	GF	NonV
Setting	-	1	3
Introductory			
Narrator	28	17	27
Dialogue	25	38	53
Developmental			
Narrator	135	109	80
Dialogue	192	281	438
Closure			
Narrator	1	-	-
Dialogue	2	-	1

3.3 Developmental text

All the sentences that develop what the introductory sentence or cluster of sentences has introduced form part of the developmental segments of the discourse. In the columns in Table 1 where figures for 'Closure' do not appear, the figures for 'developmental' include concluding sentences as well. (Unfortunately, the decision to distinguish closure from developmental was not made until the tabulation process was halfway through. As a result, I have figures for closure for only part of the material. I was convinced, however, that even such fragmentary data was better than none at all.)

Once introduced, the topic of discourse becomes referential and part of shared information. Thus, in developmental text, we can predict that there would be a preponderance of GF constructions. All across the various discourse types, this is in fact what Table 1 shows quite consistently. The differences in ratio of GF to AF sentences in each discourse type is only a matter of degree. For example, newspaper items consistently dealt with only one topic of discourse; as a result, the rest of the text after the first sentence or after the heading and subheading tended to be anaphoric. The GF sentence, being one of the anaphoric devices available in the Tagalog system (Naylor 1984), occurred much more frequently in newspaper text. Similarly, columnists tend to deal with one topic at a time and we can see that in this genre, the ratio of GF to AF sentences is quite high: roughly 2 to 1.

The figures for the play format, the television newscast, and the conversations - all conversational in character - all show remarkably high ratios of GF to AF sentences, with the conversations showing a ratio of higher than 3 to 1. (These same genres showed extremely high ratios of AF to GF sentences in introductory material.)

Again, given that AF sentences tend to code new or background information, how would they function in developmental text?

We have observed that AF sentences tended to occur in non-introductory text whenever:

(1) They coded intransitive actions. There are no other options for coding intransitives in Tagalog. Whatever their function in discourse might be, intransitive actions are coded as AF constructions. This in fact turned out to be the most common reason for the occurrence of the AF construction in developmental text;

(2) the sentence in initial position was repeated for rhetorical effect;

(3) when forming a series of parallel constructions for aesthetic effect, then an initial AF sentence would be followed by developmental AF sentences;

(4) when the focus of attention is on the actor, the AF construction may occur in developmental text. For example, in a selection on *what people do* on All Saints' Day, we observed that when talk centred on the cemetery itself and *what people did to it*, GF sentences were used; on the other hand, when talk centred on people themselves and *what they did* on that day, then AF sentences were used even in the developmental segment of the selection.

3.4 Closure

Although the figures for 'closure' are too small to be significant, it is interesting to note that they show, on the whole, a tendency to pattern similarly to introductory text.

3.5 Nonverbal sentences

As I have earlier pointed out, nonverbal clauses quite unexpectedly turned out to be of high frequency in Tagalog discourse. In Table 2, we can see that, in 7 out of 10 columns, nonverbal constructions far exceed either GF or AF constructions in relative frequency. (In my earlier study, I did not include nonverbal clauses in my tabulations. I did jot down however among my notes that with respect to a certain book, nonverbal clauses appeared to predominate.)

In the remaining three columns, the nonverbal constructions nonetheless show high frequency in two of the columns - higher than AF and not much lower than GF. Only in the television newscast was the frequency of occurrence of nonverbal sentences relatively low.

The column in Table 2 marked 'single topic' consists of articles that centre on one topic: e.g. biography, Christmas (and what it means to some famous movie stars). In this type of discourse, topicalisation in the form of nonverbal or equational sentences (pragmatically equivalent to the English cleft constructions), does occur with very high frequency.

3.6 Oral-style and written-style discourse

By 'oral-style' discourse, I refer to discourse that is conversational or conversational in style, or a simulation of conversation. It may be written or spoken text. Similarly, 'written-style' discourse refers to discourse, spoken or written, that carry most of the features of written text. Oratory, for example, may be spoken but the style is written.

The figures in all three tables show a higher ratio of GF to AF sentences in the oral-style types of discourse (the play format, the television newscast, the conversations, and the comics format). In Tables 2 and 3, we also see a remarkably high frequency of occurrence of nonverbal sentences in the oral-style discourse. These two sentence types correlate with the mode of oral communication and its setting. Unique to conversational settings is the *immediacy of reference*,[9] linguistic as well as extra-linguistic. Given the fact that GF constructions 'focus' on *what was done to something or someone that we already know about, and that nonverbal sentences are referring*, rather than narrating, predications we can only surmise that *immediacy of reference* must be a distinctive feature of spoken discourse, and by extension, of oral-style discourse. (Much work is currently being done in the area of oral versus written discourse, e.g. the work of Tannen and others in Tannen 1982, 1984.)

While these observations need - and ought - to be further explored, the figures in Tables 2 and 3 and the correlative pragmatic features of the GF and nonverbal sentence types are compelling. At least for Tagalog, they appear to be diagnostic of oral-style discourse and from the point of view of the pragmatics of focus, these figures are highly revealing of what may turn out to be one of the important functions of the GF construction: as a vehicle for immediacy of reference.

4. CONCLUSION

The relative frequency counts tabulated in Tables 1, 2, and 3 *do not reveal* a marked separation of functions-in-discourse of AF and GF constructions. The figures are equivocal and undecisive. My observations lead me to believe that the correlation between low transitivity with introductory material and high transitivity with developmental material does not hold for Tagalog, not even for narrative discourse. AF sentences (low transitivity) can code *highly active* meanings and therefore make for narrative movement. The following examples from the corpus under study will illustrate this:

(1) Umikot at humarap sa kausap ang nangangalit na tao.
turned-AF *and confronted*-AF *interlocutor the irate* linker *man*
The irate man turned and confronted his interlocutor.

(2) Sumugod ang lalaki.
chased the man
The man ran in hot pursuit.

Similarly, GF (high transitivity) constructions can code inactive, descriptive meanings, e.g. adversative examples as given above. Furthermore, GF (highly transitive) sentences, in their participial force may be descriptive in the way participial adjectives are. The incidence of AF sentences in developmental text is high; so is the incidence of GF sentences in introductory material.

It has become clear that *transitivity may not* be the relevant contrast in the pragmatics of focus in Tagalog. There are strong indications that the parameters within which the verbal system of Tagalog functions are marked by other contrasts; e.g. active-stative, centrifugal-centripetal, and others that we have yet to discover.

We have begun to ask if in fact discourse coherence is itself a function of the verbal system in Tagalog. The fact that so very many nonverbal sentences occur in recorded usage has brought up the question of how do verbal and nonverbal constructions function together in the organisation of discourse.

We have begun to glean some likely answers. Nonverbal clauses must also contribute to the flow of discourse. With their occurrence in great numbers, discourse would come to a long standstill if nonverbal clauses contributed nothing to the flow of discourse. The fact is that *all clauses* - verbal and nonverbal - *have a topic*. Nonverbal sentences are not endowed with transitivity. Furthermore, we know that the focus relationship singles out the topic *ang*-NP, thus making the topic salient. Perhaps, ultimately, focus functions as a vehicle for *topic salience*, and in the final analysis, focus functions as a coherence device only by virtue of its association with the topic.

From the lack of sharp definition of discourse functions of the AF and GF clause types, from the remarkably high incidence of nonverbal sentences, and from all the other foregoing observations, we are led to the conclusion that Tagalog discourse is organised in terms of *topic - topic continuity* and *topic movement* rather than transitivity (cf. McGinn 1982).

Any account of Tagalog discourse organisation in terms of transitivity and the function of verbal clause types in the organisation of discourse ignores a *large part* of discourse as well as an *integral part* of the system. Transitivity fits Tagalog discourse only loosely and incompletely; it leaves out much that needs to be accounted for if we are to describe the discourse dynamics of Tagalog with a modicum of adequacy.[10]

While this study may prove sufficient for the study of the pragmatics of focus, the study of the pragmatics of focus is not sufficient for a study of the dynamics of Tagalog discourse.

This study, however, is a step in the right direction. We have already begun to take the next step - in the same direction.[11]

NOTES

1. In Philippine linguistics, the term *focus* has a unique meaning. It refers to the syntactic-semantic relationship between the verb and the surface subject, signalled by the verb's focus affix in conjunction with the subject form of noun phrases and pronouns. For example, a sentence is in *actor focus* if the surface subject is in the role of actor and the verb carries an actor-focus affix.

2. See Naylor 1984.

3. M.A.K. Halliday (1973, personal communication): Context does not determine choice and it is still the speaker who *chooses*; context merely constrains.

4. We must caution here that it is important not to confuse high frequency of occurrence with structural basicness. Markedness, rather than basicness, is what is correlatable with frequency.

5. The modification structure is: Head + Linker na + Adjective/Adverb. The so-called relative-clause structure is: Head + Linker na + Verb. Thus, ang titser na maganda *the beautiful teacher* is no different from ang titser na umalis *the teacher who left* (lit. **the left teacher*). In fact, the underlying predications are structurally identical as well in Tagalog: maganda ang titser *the teacher (is) beautiful* and umalis ang titser *the teacher left*, respectively.

   If I understand Stanley Starosta correctly, he shares this general point of view (personal communication at FOCAL). Barry Miller, in his ECAL-IV paper (August 1985) also concurs. In Naylor 1976, this has been argued within a theory of attributive syntax in Tagalog.

6. Dr Mary Bresnahan had made a similar observation in the course of writing her dissertation, based on a Tagalog novel and other Tagalog texts.

7. A more detailed account of the discourse function of nonverbal sentences is the subject of another study, now in progress, part of which is reported in Naylor 1985.

8. It must be kept in mind that we are talking about the high frequency of AF sentences in dialogues and conversations in *introductory material only*. As we shall see below, the AF sentence is *not* the most frequently occurring sentence type in these genres in terms of overall frequency.

9. In more recent work that touches on anaphora, we observed that in oral-style discourse, the constituent that is subject to anaphoric processes (substitution and deletion) is never far - usually, only one clause away but never more than two clauses away, either immediately preceding or the next but one clause before. In this connection, we can perhaps talk about 'phoric distance', which in oral-style discourse has to be short but which, in written-style discourse may be longer, due to the nature of the setting. (Since this is not the subject of the present study, I cannot go into further detail on this.) It may be that GF and nonverbal sentences are diagnostic of oral-style discourse because of their association with short 'phoric distance'. These observations are still highly tentative at this time, however.

10. cf. Naylor 1986. In most recent developments of my work on focus, I have come to the conclusion that focus and transitivity are systems that are entirely different but interrelated in ways that parallel the contrast between aspect and tense. Focus is to aspect as transitivity is to tense. Just as languages in general are said to have both tense and aspect, so do they have both transitivity and focus. But just as some languages are said to have verbal systems based on aspect and others on tense, we can also say that some languages have verbal systems based on focus and others on transitivity. Focus is perspectival and therefore pragmatically based. Transitivity, on the other hand, is a matter of case relations and therefore semantically based. Both transitivity and focus are, however, realised by the syntax. (The parallels with tense and aspect are fairly obvious. Like tense and aspect, focus and transitivity are interrelated and they perform similar functions in the grammar of different languages but they are nonetheless two entirely different systems based on contrasts that are quite different in nature.)

11. cf. Naylor 1985. Other papers given at the Fourth Eastern Conference on Austronesian Languages (ECAL-IV), held in Ann Arbor, August 2-4, 1985 showed a similar orientation: e.g. those of Miller, Weeda, Basham, etc.

REFERENCES

CEÑA, Resty

1977 The 'passive' as basic sentence in Tagalog. Abstract of paper presented to the LSA Winter Meeting, Chicago.

DARDJOWIDJOJO, Soenjono

1977 The semantic structure of the adversative *ke--an* verbs in Indonesian. Paper given at the Austronesian Symposium, LSA Linguistic Institute, August 18-20. Honolulu.

HALLIDAY, M.A.K.

1968 Notes on transitivity and theme in English. *Journal of Linguistics* 4:179-215.

1973 *Explorations in the functions of language*. London: Edward Arnold.

HALLIDAY, M.A.K. and Ruqaiya HASAN

1973 *Cohesion in English*. (Pre-publication MS, subsequently published 1976, London: Longman.)

HOPPER, Paul J. and Sandra A. THOMPSON

1980 Transitivity in grammar and discourse. *Language* 56/2:251-299.

KESS, Joseph F.

1979 A psycholinguistic frame of reference for focus and topic in Philippine languages. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, Fascicle 1:443-461. *PL*, C-61.

McCUNE, Keith

1979 Passive function and the Indonesian passive. *Oceanic Linguistics* 18/2:119-170.

McGINN, Richard

1982 A principle of text coherence in Indonesian languages. Paper presented to AAS Meeting, Chicago, April 4. MS.

MILNER, George B.

1973 It is aspect (not voice) that is marked in Samoan. *Oceanic Linguistics* 12:621-639.

NAYLOR, Paz Buenaventura

1973 Topic, focus, and emphasis in the Tagalog verbal clause. Ph.D. dissertation, University of Michigan. Ann Arbor: University Microfilms.

1974 On contextual aspects of topicalization. Paper presented to FICCAL, January 2-4, Honolulu.

1975 Topic, focus, and emphasis in the Tagalog verbal clause. *Oceanic Linguistics* 14:12-79.

1976 Linking, relation-marking, and Tagalog syntax. Paper presented at SECAL, May 2-4, Ann Arbor (subsequently published in Naylor, ed., 1980, *Austronesian studies: Papers from the Second Eastern Conference on Austronesian Languages*, 33-50. Michigan Papers on South and Southeast Asia CSSEAS Publications, The University of Michigan.

1984 When must a definite semantic object be the syntactic subject in Tagalog? Paper presented at the Fifth Groningen Roundtable, June 15-19.

1985 On the centrality of topic in the structure of Tagalog discourse. Paper presented at ECAL-IV, August 2-4, Ann Arbor.

1986 Focus *versus* transitivity. Paper presented at the University of Michigan Linguistics Colloquium. March 27, 1986.

PIKE, Kenneth L.

1963 A syntactic paradigm. *Language* 39:216-230.

1964 Discourse analysis and tagmeme matrices. *Oceanic Linguistics* 3:5-25.

TANNEN, Deborah

1982 Oral and literate strategies in spoken and written narratives. *Language* 58/1:1-21.

TANNEN, Deborah, ed.

1984 *Coherence in spoken and written discourse*. Norwood, N.J.: Ablex.

TCHEKHOFF, Claude

1973 Verbal aspects in an ergative construction: an example in Tongan. *Oceanic Linguistics* 12-607-620.

SOME CONSEQUENCES OF CAUSATIVE CLAUSE UNION IN TAGALOG

Videa P. De Guzman

1. INTRODUCTION

Two previous works dealing with causative constructions of Philippine languages have shown the manifestation of what is known in Relational Grammar (RG) as Causative Clause Union (CCU). The first study by Bell and Perlmutter (1981) titled "Causative Clause Union and Advancements in three Philippine languages" is cast in the passive analysis (PA) while the other by Gerdts (1983), "Antipassives and causatives in Ilokano: evidence for an ergative analysis" employs the ergative analysis (EA). Interestingly enough, both approaches, each with its own accompanying device, are able to show how CCU accounts for the behaviour of grammatical relations in causative constructions with biclausal structure representation. The present paper applies both accounts to Tagalog and analyses the consequences of each account. Following the arguments presented in support of each account, it will show their shortcomings. As well, it will suggest that there are variations in case and voice marking that are not exclusively syntactically motivated. On the contrary, certain case and voice marking rules depend on the semantic orientation of the verb and causative verbs manifest this in their treatment of the two complement nuclear terms.

Based on Postal and Perlmutter's proposal, CCU makes the universal prediction that the grammatical relations (GR's) borne in the matrix clause by the final nuclear terms of the complement are as follows: the final subject of an intransitive complement and the final direct object of a transitive complement are the direct object of the matrix clause, and the final subject of a transitive complement is the indirect object of the matrix clause (Bell and Perlmutter 1981:3). Put another way, the downstairs final Abs(olutive) is upstairs object or term 2 and the downstairs final Erg(ative) is upstairs indirect object or term 3. Prior to an analysis of causative constructions, it will be helpful to review the case and voice marking rules in Tagalog which parallel those stated by Bell and Perlmutter for simple clauses because they claim that the same rules apply to complex clauses. A nominal heading a final 1-arc in the highest clause in which it heads a central relation arc must be in the Nom(inative) case; one that has a final 2-arc must be in the Acc(usative); one that heads a final 3-arc or a final Loc-arc, in the Obl(ique); one that heads a final Ins-arc must be in the Gen(itive). A 1-chômeur must be in the Gen(itive) and one that heads a final Ben-arc is in the Obl case but it is introduced by the preposition para *for*. The voice marking rules may also be simply stated as follows: When the final 1 is not a successor of any other relation, the verb must be in the Active voice. If the final 1 is the successor of 2, then the verb must be in the Obj(ective) voice; if 1 succeeds 3 or Loc, then the verb is in the Ref(erential) voice. It

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 59-72. *Pacific Linguistics*, C-93, 1986.

is in the Ins(trumental) voice if the final 1 is the successor of a grammatical relation other than those mentioned in the preceding.

2. PA OF CAUSATIVE CONSTRUCTIONS

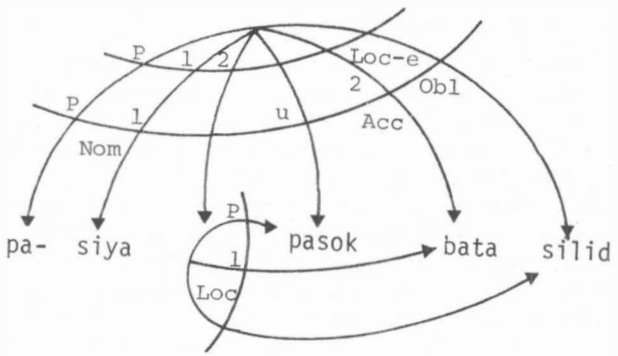
Initially, a causative structure has a matrix clause and a complement transitive or intransitive clause. In Tagalog the matrix or upstairs clause has a P(redicate) of causation indicated by the affix pa- and two nuclear terms 1 and 2. The complement or downstairs clause heads the 2-arc. Sentences (1) and (2) show an intransitive and a transitive complement clause, respectively:

(1) nagpapasok siya ng bata sa silid
Act-caus *enter* Nom *she* Acc *child* Obl *room*
She made a child enter the room.

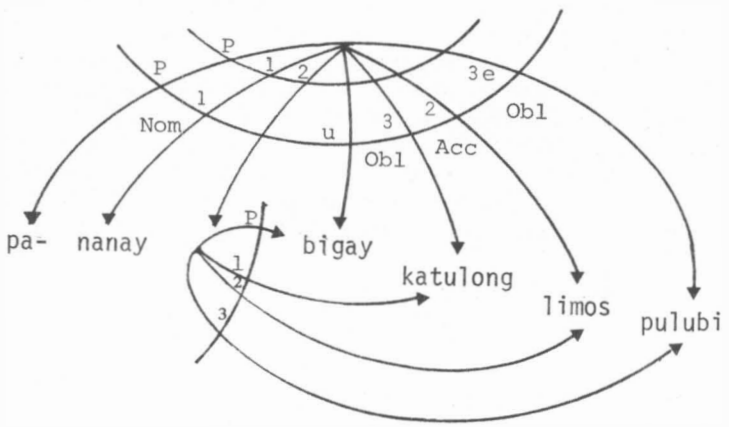
(2) nagpabigay ang nanay sa katulong ng limos sa pulubi
Act-caus *give* NOM *mother* Obl *maid* Acc *alms* Obl *beggar*
Mother had the maid give alms to the beggar.

As stated earlier, CCU accounts for the GR's of the nominals in the complement clause as they become part of or united with the matrix clause. The network diagrams showing CCU for (1) and (2) are as follows:

(1')



(2')



When the nuclear terms which unite in the matrix clause in (2') each end up as a final 1 via the characteristic advancement rules, we get two voice markers that differ from what the voice marking rules predict. Sentences (3) and (4) show deviations but not (5):

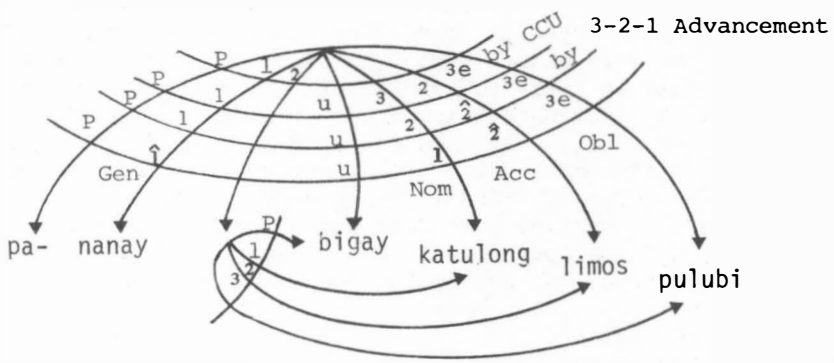
(3) papagbibigay<u>in</u> ng nanay <u>ang katulong</u> ng limos sa pulubi
Obj-caus *give* Gen *mother* Nom *maid* Acc *alms* Obl *beggar*
Mother will have <u>the maid</u> give alms to the beggar.

(4) <u>i</u>pabibigay ng nanay sa katulong <u>ang limos</u> sa pulubi
Ins-caus *give* Gen *mother* Obl *maid* Nom *alms* Obl *beggar*

(5) pabibigy<u>an</u> ng nanay sa katulong ng limos <u>ang pulubi</u>
Ref-caus *give* Gen *mother* Obl *maid* Gen *alms* Nom *beggar*

In (3) where we find the former term 3 as final 1, the voice marking on its co-occurring verb is not the Referential -an as stipulated in the rules when a 3-1 Advancement takes place. Instead, the verb is in the Objective voice with the affix -in. To mark it with -an will render the sentence ungrammatical. If we assume, however, that this former 3 is a 2, so that a 2-1 Advancement accounts for the Objective voice marking on the verb, the conflict will be in the case marking of this term. As a final 2, it is not marked with the Accusative ng (pronounced naŋ) according to the rules, but with the Oblique sa. To use the former case marker will result in an ungrammatical structure.

Bell and Perlmutter propose to remedy this conflict by positing an obligatory rule labelled 3-2-1 Advancement rule which not only insures the correct Oblique case marking of a final 3 but also the proper Objective voice marking on the verb when this 3 ends up as a final 1. Thus, (3) has the following network diagram:

(3') 

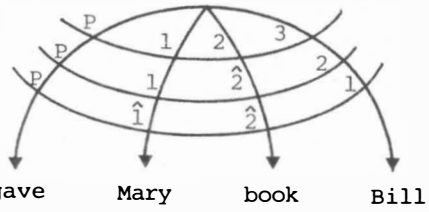
In (4) it is term 2 that is succeeded by final 1 as shown by its Nom case marking. But contrary to the voice marking rule stated previously, its co-occurring verb is not in the Objective voice.[1] Rather, it is in the Instrumental voice. Finally, with 3e(meritus) ending up as final 1 as shown in (5), its co-occurring verb is in the Referential voice as though 3e were no different from an ordinary 3 undergoing a 3-1 Advancement.

3. ANALYSIS OF THE PA ACCOUNT: CCU AND 3-2-1 ADVANCEMENT

This brings us to an analysis of two consequences brought about by CCU as revealed in sentences (3) and (4). The first consequence has to do with positing the companion rule 3-2-1 Advancement, and the other, with the Instrumental voice marking of the verb when its complement 2 becomes final 1.

3.1 The 3-2-1 Advancement rule

There is something to be said about the formulation of 3-2-1 Advancement and the constraint on its application. We are familiar with two individual advancement rules applying in succession such as 3-2 and 2-1 as in the English example: Bill was given a book by Mary. The network corresponding to this structure is as follows:



Unlike this illustration, 3-2-1 is just one rule consisting of an obligatory series of transitions. This type of rule has been defended by Bell and Perlmutter (1981:41-48) showing that it is not necessary for the intermediate transition to have a corresponding manifestation as a final stratum and that this type of device also exists in languages as diverse as Quiché, French and Chicewa.[2] Due to limitations of time and space, we will forego the evidence they provide. Suffice it to stress that the rule allows term 3, predicted by CCU and which is correctly marked with the Oblique case, to take a verb in the Objective voice when this 3 ends up as a final 1 by undergoing an intermediate transition to 2. Comparing the network for the English sentence and that of (3), however, we won't be able to detect that two different types of rules account for each of the identical representations in terms of transitions undergone. Moreover, this type of two-stage obligatory transition rule does not stipulate its restrictions. If we consider the following sentences, we will note that the forms of the verbs in the (b) sentences are not predictable from the voice marking rules:

(6) a. naglaba siya ng damit
Act-*washed* Nom *he* Acc *dress/clothes*
He washed (some) clothes.

b. nilabhan niya ang damit
Ref-*washed* Gen *he* Nom *clothes*
He washed the clothes.

(7) a. magtatapon siya ng basura
Act-*will throw away* Nom *he* Acc *garbage*
He will throw away some garbage.

b. itatapon niya ang basura
Ins-*will throw away* Gen *he* Nom *garbage*
He will throw away the garbage.

With a 2-1 Advancement, the verb in (6b) is marked with the Referential voice and that in (7b), with the Instrumental voice. Following the same motivation for positing 3-2-1 Advancement, we may ask if there is a constraint on positing a 2-3-1 rule for (6b) and a 2-Ins-1 for (7b). Apparently, this has to be adopted if we want to avoid subcategorising verbs according to the various affixes they take when 2 becomes final 1. So far, RG has made no provision for such distinctions except to label them 'irregularities'. It is, therefore, still unclear what the 3-2-1 type of rule precisely means and what the conditions for formulating it must be.

Even if we accept the validity of 3-2-1 in form and content, it still raises the question of generality of application. It only applies to this particular term 3 of causatives which originates from complement 1 of a transitive clause. It does not apply to 3's of non-causative verbs nor to 3e's of causatives. What this rule suggests is that this term 3 is different from either 3's or 2's of non-causative verbs. And this is evidently the reason why it requires a different type of advancement rule, i.e. to serve this unique GR. Granted further that we accept this rule of 3-2-1, we will find that with certain classes of transitive verbs, a 3-1 Advancement is the only appropriate rule for the construction in question. To illustrate:

(8) pakikitaan ko ang nanay ng pelikula
Ref-caus-*see* Gen *I* Nom *mother* Acc *film*
I will have mother see a film/I will show mother a film.

The unfortunate implication of this counter-example is that 3-2-1 applies to 3's of certain classes of causative verbs, while 3-1 applies to certain others, thus further diminishing the generality of application of 3-2-1 Advancement.

3.2 The Instrumental voice and 2-1

Bell and Perlmutter claim that when term 2 in the matrix clause which comes from the complement 2 is taken over by final 1, the verb it takes is in the Objective voice (see note 1). Although this observation may be true of a certain derived class of verbs, e.g. MA-abilitative, resulting in the ambiguous Objective voice forms, the general tendency is to distinguish the voice forms of the verb that indicates a final 1 as the successor of a term 3 from that of a term 2. Sentence (4) shows that in Tagalog this particular 2-1 Advancement, i.e. matrix 2 from complement 2, is marked with the Instrumental voice (with the Ins affix i-), just as Bell and Perlmutter observed in Ivatan (1981:67).[3] Evidently, this is another adverse consequence of CCU because it runs counter to the prediction of the rule which states that an Objective voice with the affix -in marks the verb when 2-1 occurs. Similar to term 3 of causatives, discussed in the preceding section, term 2 from complement 2 is different from other term 2's of both causative (from complement 1 of an intransitive clause) and non-causative verbs. In the manner of the 3-2-1 device, it may be suggested that there must be a corresponding 2-Ins-1 rule which will trigger the appropriate voice marking here. This possibility is not even considered by Bell and Perlmutter to account for the same phenomenon in Ivatan, so we can only surmise that there may be an implicit restriction that only nuclear terms may appear in intermediate transitions to conform to the established hierarchy of terms. The only alternative solution which seemed satisfactory to them was to posit an irregular voice marking rule that serves this specific requirement (1981:67-70). This rule which they

identify as a language specific rule, unfortunately, misses the significant generalisation that may be captured in the structure involved across Philippine languages.

To point out one more complication, there is a class of verbs in Tagalog such as those exemplified in (6) that are marked with the Referential voice affix -an when the same complement 2 is final 1 in the matrix clause. For example:

(9) palalabhan ng nanay sa katulong ang damit
Ref-caus-*wash* Gen *mother* Obl *maid* Nom *dress*
Mother will have the maid wash the dress.

Again, we may ask whether a 2-3-1, instead of a 2-Ins-1, would be allowable in order to trigger the appropriate voice marking.

Judging from the fact that two special rules have to be formulated to apply obligatorily and exclusively to term 3 (from complement 1) and to term 2 (from complement 2) when they end up as final 1, there is reason to believe that these two rules for causative constructions are in effect implying the necessary modification that has to be built into the voice marking rules stated for simple clauses. This modification is well motivated in view of having to accommodate one more GR, the complement 1 of transitives, when all other nuclear terms are already occupied. As Comrie (1976:261) discusses, one strategy for accommodating this extra noun phrase is by means of doubling up in one of the syntactic positions of the sentence. In Tagalog, he identifies doubling on indirect object (1976:279;310) and preferably, if one of them is changed by focusing, i.e. becoming final 1. What he did not mention is that when this complement 1 appears as a final 1, thereby avoiding doubling on term 3, it takes an Objective voice verb. On this basis, we can say that the doubling occurs on direct objects since in the verbal paradigm we seem to have two Objective voices, one marked -in for the former complement 1 and another marked i- for the former complement 2. Yet, because we have to refer to its original termhood in the complement clause to identify its case and voice markers, it is no more economical and adequate to account for this complement 1 in question as a different GR which shares the same case marking feature with term 3 but the same voice marking feature with term 2. It remains distinct from either 3 or 2 and, thus, it does not violate the Stratal Uniqueness Law when it co-occurs with either of them.

3.3 On the case and voice marking rules

Some observations pertaining to the case and voice marking rules used with simple clauses have to be mentioned in considering Bell and Perlmutter's conclusion that the same rules are also used with causatives (1981:53). Going back to the case-marking rules stated earlier (see p.59), a final 2 can only be marked in the Accusative, i.e. ng marker or its equivalent substitute forms. However, in more recent studies of Tagalog (Ramos 1974:100-101;130; McFarland 1976:6-7; De Guzman 1978:75-79), it has been recognised that direct objects or final 2's are manifested only in the Oblique form when they are either proper nouns or personal pronouns. For example:

(10) naghintay ang nanay {kay Maria/*ni Maria (Obl / Acc); sa kaniya/*niya (Obl *her* / Acc *her*)}
Act-*waited* Nom *mother*
Mother waited for {*Maria*, *her*}.

Verbs identified as extension verbs require their co-occurring objects to be marked in the Oblique case and not in the Accusative, as in the following:

(11) tumulong siya sa mga nasunugan
Act-*helped* Nom *he* Obl pl. *fire-victim*
He helped the fire-victims.

(12) siya ang humalik sa bata
Nom *he* nomlzr Act-*kissed* Obl *child*
It was he who kissed the child.

Moreover, certain classes of verbs allow their objects to be marked with either the Accusative or the Oblique case to distinguish the meaning indefinite/non-specific or definite/specific, respectively. For example:

(13) nagbantay ako ng/sa bata
Act-*looked after* Nom *I* Acc/Obl *child*
I looked after a/the child.

Although it occurs occasionally in the basic active voice constructions, the Oblique-marked final 2 is more commonly found in cleft and in non-active constructions.

In causative constructions, final 2's from complement 1 of intransitives usually allow the same case marking alternation as in the following:

(14) nagpatulog ako ng bata
Act-caus-*sleep* Nom *I* Acc *child*
I made a child sleep.

(15) ako ang nagpatulog ng/sa bata
Nom *I* nomlzr Act-caus-*sleep* Acc/Obl *child*
It was I who made a/the child sleep.

Final 2's from complement 2 with active verbs exhibit the same case marking alternation in cleft constructions.

What has been shown above is that a nominal marked with the Oblique may not necessarily be a 3; it may be a definite or a specifiable 2. If this avenue is taken as valid and complement 1 is considered a term 2 in the matrix, regardless of the transitivity of the downstairs clause, the need for the 3-2-1 rule vanishes. In its place, however, a supplementary rule on case marking the complement 1 from transitives as an Oblique 2 upstairs has to be incorporated.

Given the voice marking rules (see p.59), the Objective voice is associated only with the affix -in. This is not entirely accurate because previous works on Tagalog grammar have identified i- and -an as the other affixes that mark other semantic classes of verbs in the Objective voice. In fact, other classes

are unmarked. It has been recognised that patterns of voice affixes, primarily in the Active and Objective, correspond to semantic groupings of verbs (Schachter and Otanes 1972; Ramos 1974; McFarland 1976; De Guzman 1978). The Objective voice affix -in generally corresponds to the Active voice affix -um- manifested by verbs that are labelled either [+change of state] or [+action toward the agent]. An equally large number of verbs (and in fact more, by McFarland's survey) are marked with i- in the Objective and mag- in the Active. These verbs usually involve objects being transported or changed in position. Lastly, the class of Objective voice verbs that take the affix -an indicate a surface change and they also take mag- in the corresponding Active voice. This variance suggests that the voice marking rules as stated earlier which make exclusive reference to GR's need some further proviso to trigger the appropriate voice affix marker. It is important to formulate these rules on the basis of the semantic subclassification of the verbs because the voice marking in the causative constructions makes reference to the same subclassification. For example, the non-causative verbs that are marked -in and i- in the Objective voice are all marked regularly with i- in the causative when the complement 2 becomes final 1; but those that are marked with -an continue to be marked in the causative form by the same affix. Thus, bilhin *to buy* vs. ipabili, ibigay *to give* vs. ipabigay, labhan *to launder* vs. palabhan. Similarly, the parallel semantic distinction made in the Active voice is carried over to the morphological structure of the causative verb stem in the Objective voice, when the complement 1 becomes final 1. Although the voice affix is regularly -in (with psychological verbs taking -an), the stem of the causative verbs that correspond to Active -um- verbs takes a verb root as base whereas that which corresponds to mag- verbs takes a pag- stem. For example pa+kain+in from kumain *to eat* vs. pa+pagbigay+in from magbigay *to give*. The complexity and variation in the morphological structure of causative verb stems cannot be explained by either CCU or the advancement rules in PA because these are not syntactic problems. But there is no doubt that their semantic and morphological features interact with the syntactic rules in question.

Except for the two nominals we have been examining, complement 1 and complement 2 of transitives, all other co-occurring complement nominals bear the same grammatical relations to the causative verb. Thus, the case and voice marking rules that apply to them in simple clauses also apply in the causative constructions. Take sentence (5) for example. Here, the former complement 3 (indirect object) which becomes a 3e(meritus) upstairs undergoes a simple 3e-1 Advancement. Its verb is marked with the Referential voice -an, and when it is a final 3e as in (4), it is case-marked Oblique. Actually, whether it is labelled 'emeritus' or not its grammatical status remains the same. Yet, this identification has to be devised in order to distinguish this 3 (from complement 3) and the 3 that originates from complement 1. It will be noted, however, that the standard voice marking rule must further incorporate 3e in the 3-1 rule. Now, if all other relations take the same case and voice marking in both simple and causative clauses, it can only mean that their GR's are not affected by CCU.

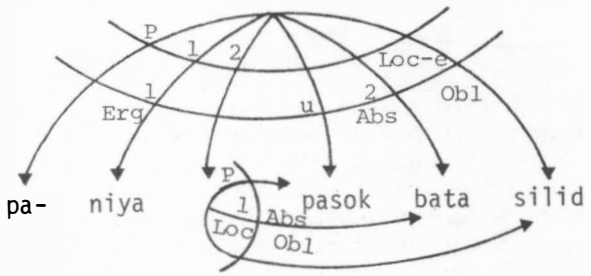
In sum, we can agree with one part of CCU's prediction that complement 1 of an intransitive clause becomes term 2 in the matrix. But it has to allow for an alternate Oblique case with certain verbs and/or in certain constructions. As for the other portion of CCU which designates complements 1 and 2 of a transitive clause as 3 and 2, respectively, there appears to be a viable alternative as will be proposed later.

4. EA OF CAUSATIVE CONSTRUCTIONS

Let us now turn to Gerdts' (1983) analysis of the causative constructions from an ergative view. She shows that in these constructions both CCU and anti-passive (AP) are manifested. By Postal and Perlmutter's CCU rule, complement final Abs is upstairs 2 and complement final Erg is upstairs 3. Corresponding to the basic sentences (1) and (2) in PA (see p.60), the following are the basic ones in EA with their network diagrams:

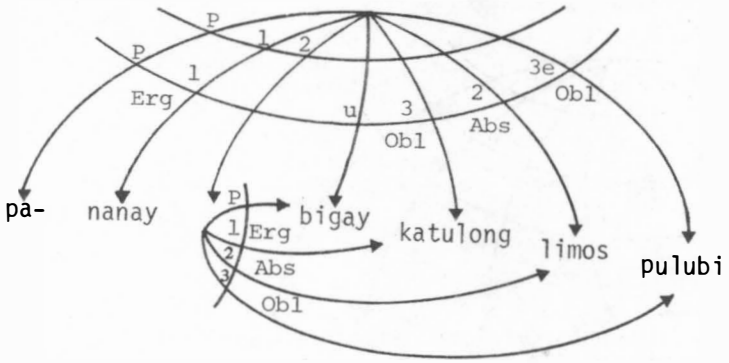
(1a) papapasukin niya ang bata sa silid
Obj-caus *enter* Erg *she* Abs *child* Obl *room*
She will make the child enter the room.

(1a')



(2a) ipabibigay ng nanay sa katulong ang limos sa pulubi
Ins-caus-*give* Erg *mother* Obl *maid* Abs *alms* Obl *beggar*
Mother will have the maid give the alms to the beggar.

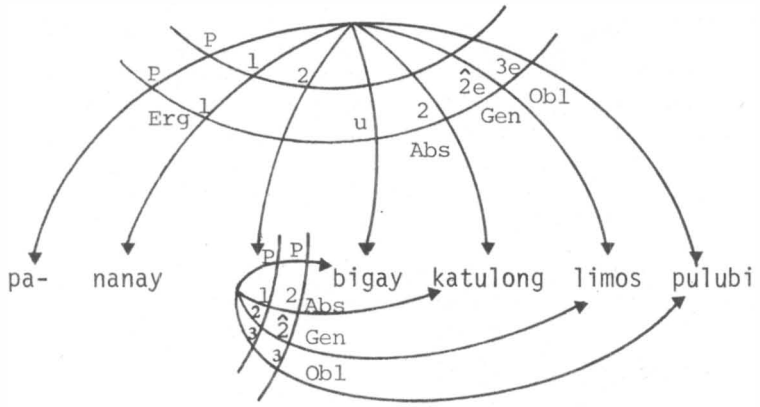
(2a')



In (1a) where the complement clause is intransitive, the complement final Abs which is upstairs 2 is properly marked with the Objective voice. On the other hand, when the same final Abs complement from a transitive clause becomes final 2 upstairs its verb is marked with the Instrumental voice. The Objective voice form as given in (3), repeated below as (3a), has the complement Erg as the final Abs in the matrix clause. For such a clause, the necessary rule is 3-2 Advancement, but which according to the voice marking rules should yield a verb in the Referential voice. Gerdts' solution to this predicament is to apply the AP rule in the complement clause before effecting CCU as follows:

(3a) papagbibigayin ng nanay ang katulong ng limos sa pulubi
Obj-caus-*give* Erg *mother* Abs *maid* Gen *alms* Obl *beggar*
Mother will have the maid give alms to the beggar.

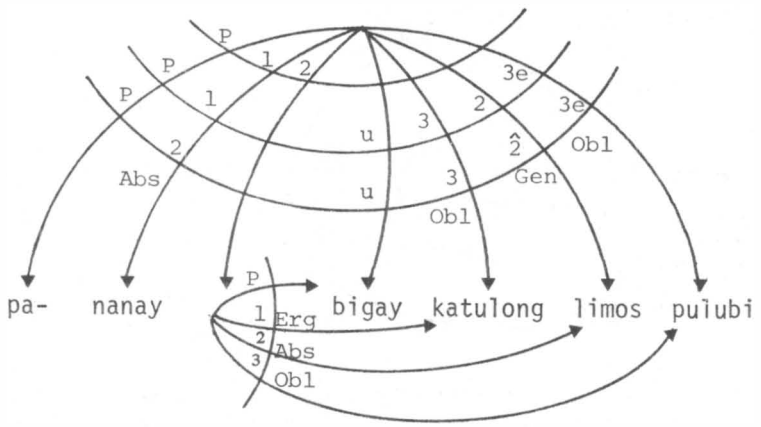
(3a')



AP turns complement 1 to 2 making the former 2, a 2-chômeur. By CCU, final 2/Abs downstairs becomes 2 upstairs and 2-chômeur becomes 2-chômeur-emeritus. (As an inherited GR or an emeritus, it is marked with the same case as a 2-chômeur which is the Genitive.) One strong argument for this account is that it does not need any other special mechanism, such as the 3-2-1 Advancement of PA, to trigger the correct voice and case marking. Notice that the effect of AP is to detransitivise the clause. Thus, sentence (2) repeated below as (4a) is shown to employ AP this time in the matrix clause.

(4a) nagpabigay ang nanay sa katulong ng limos sa pulubi
Act-caus *gave* Abs *mother* Obl *maid* Gen *alms* Obl *beggar*
Mother had the maid give alms to the beggar.

(4a')



In the same way that AP is employed in simple transitive clauses, it can also be applied to causative constructions. By definition, a transitive clause has to have both a final Erg and a final Abs, but since AP changes the Erg to Abs, the resulting structure is intransitive. In the antipassive construction, the verb is in the active voice and the initial 2 ends as a 2-chômeur which is case-marked Genitive.

One other advantage Gerdts claims for AP is that it explains the affixation of pag- in the verb stem. As we can observe in (4a) and (3a), pag- (or nag-) is affixed before the stem pabigay and the root bigay, respectively, supposedly as the morphological effect of AP. The affix pa- gets attached as a consequence

of CCU. According to Gerdts, the order of the affixes pag- after pa- in the verb stem build-up results from the application of CCU first and then AP next.

5. ANALYSIS OF THE EA ACCOUNT

From the above examples, it appears that between AP in EA and 3-2-1 Advancement in PA, the former is preferable as a companion rule to CCU both for its generality of application and for the explanation it provides for the verb morphology. These two advantages, when analysed more closely, leave us with some provoking questions. Firstly, not all verbs are affixed with pag- when AP applies to the basic transitive structure. Other active voice verbs, the so-called UM-verbs, do not exhibit this affix. For example: kumain *to eat*, humiram *to borrow*, kumuha *to get*, bumuli *to buy*, tumanggap *to receive*, etc. Primarily, all active verbs which may be identified as actions toward the agent or actions internal will be ungrammatical with the affix pag-. In causative constructions, the same verb stems which belong to the UM-class are not marked by pag- when AP applies to the complement clause. The morphological side effects of AP have to be modified then for Tagalog (and even for Ilokano and the other Philippine languages that make the UM- and MAG- distinction) to account for the correct active voice affixation. It will be instructive to remember that the non-causative verbs in the Objective voice marked with -in, as previously mentioned, are marked with -um- in the active voice; those that are marked with i- or -an correspond to mag- forms in the active voice. Obviously, this patterning of affixes cannot be accounted for syntactically. They are either morphologically or semantically bound.

Another observation which has some theoretical implications for the application of AP may be illustrated in the following examples:

(16) ipagbibigay ko ang nanay ng abuloy sa Cancer Society
Ins-*give* Erg *I* Abs *mother* Gen *contribution* Obl
I will give a contribution to the Cancer Society for mother.

(17) ipagpapaluto ko sa katulong ang nanay ng pagkain
Ins-caus *cook* Erg *I* Obl *maid* Abs *mother* Gen *food*
I will have the maid cook some food for mother.

As indicated by the nominal marked Abs, these two sentences have undergone the Ben-2 Advancement rule; the Ben nominal is marked Abs and the verb is in the Ins voice. Both verbs show the presence of the affix pag- which is presumably introduced by AP. But contrary to the effect of AP that the Erg nominal be succeeded by a final Abs, it is in the Obl case. Two questions may be raised in this connection as follows:

(a) Is the affixation of pag- to the verb stem also a side effect of Ben-2 Advancement rule, just like the AP rule? If so, how will the rule be constrained so that no pag- is affixed to UM-verbs when the rule applies?

(b) Is there a relation existing between the Active voice and the Benefactive voice (marked by the Ins voice affix) which would explain the identical verb stem forms that they take, i.e. verbs belonging to the UM-class are not marked with the affix pag- in either voice; those that belong to the MAG-class are?

One other problem we detect in the EA account pertains again to the complement 2. In EA, this complement is the Abs downstairs and by CCU, it remains to be 2 in the matrix clause. As a final 2 or final Abs, the corresponding voice marking rules for simple clauses will mark its co-occurring verb incorrectly with the Objective voice affix -in, instead of the appropriate affix i-. It appears that this situation can only be remedied by formulating an appropriate voice marking rule. For this particular final Abs to co-occur with a verb in the Ins voice, it should either be an Instrumental or a non-term before becoming a final 2 Abs. EA's recourse may be a 2-Ins-2 rule, in the manner of PA's 3-2-1. Conceptually, however, this rule is indisputably odd.

6. SUMMARY AND CONCLUSION

From the two approaches we have analysed, the following are the features characteristic of Tagalog causative constructions:

(a) Complement 1/Abs from a final intransitive clause and complement 1/Erg from a final transitive clause, when they end up as a final 1 (Nom in PA) or a final 2 (Abs in EA) after CCU, take a verb in the Objective voice. As non-final Nom/Abs, these two complements behave differently in that the former is an upstairs 2 whereas the latter is an upstairs 3. Accordingly, they are marked with the Accusative/Genitive and the Oblique, respectively.

(b) Complement 2 from a transitive clause behaves differently from the two complements above in that when it is a final 1 (in PA) or a final 2 (in EA) in the matrix clause, its verb takes the Instrumental voice, not the Objective voice as predicted by the voice marking rules.

While both PA and EA, accompanied by 3-2-1 and AP, respectively, claim to be able to account for (a), regardless of their undesirable consequences, neither approach provides a solution to the problem identified in (b). As they stand, each is not a compelling account. In fact, if we consider the two complements in (a) above which are united by the Objective voice marking of their co-occurring verb when they are final Nom/Abs in the matrix clause, there is reason to believe that they may be actually 2's upstairs. This consideration is in line with Gibson's proposal of a CCU II (1980) which states that a nominal heading a final 1-arc in the complement heads a 2-arc in the matrix clause, regardless of the transitivity of the complement clause. Following EA and making use of AP only to advance matrix Erg to Abs, we would need a 2-3 retreat rule to mark the former (downstairs) Erg with the Oblique case, after it unites upstairs as a 2, when a different complement is taken over by the final Abs. Moreover, complement 2 identified in (b) becomes a 2e(meritus), and as such needs a rule to mark it with the Genitive case; this non-term relation like other emeritus relations may assume a final Abs relation in which case a rule must mark its verb in the Instrumental voice.

In view of the possible alternative accounts for Tagalog causative constructions which RG follows, the claim of one version is correspondingly weakened by the existence of another version. As shown by the different consequences of adopting CCU (or even CCU II), there must be established some tighter constraints on the form of the companion rules of CCU and a requirement on the generality of their application. It is not enough for the account to generate an air of systematic rigour, but, more importantly, it should provide an adequate explanation for syntactic similarities and distinctions. Granted that these criteria for a desirable account are met, the one that explains the non-isomorphic relation

between nuclear terms and case/voice markers existing in a language such as Tagalog and the other Philippine languages, which others have claimed as emanating from semantic distinctions is to be preferred.

NOTES

1. The Hiligaynon example used by Bell and Perlmutter (1981:26) from which they concluded that the 2-1 in question is in the Objective voice comes from a different class of verbs, the MA-class. The ordinary form of mapaluto, which they used, is in fact ipaluto with the Ins affix i-.
2. This position is criticised by W.D. O'Grady in his 1980 article.
3. The other major Philippine languages, except Kapampangan, likewise mark this particular 2-1 advancement with the Instrumental voice. In Maranao, although both complement 1 and complement 2 take the Objective voice affix -en when they become final 1, the two verb forms remain distinct because of the difference in their stems, e.g. pakatabasen for the former and pakitabasen for the latter.

REFERENCES

BELL, Sarah J. and David PERLMUTTER

1981 Causative clause union and advancements in three Philippine languages. (MS).

COMRIE, Bernard

1976 The syntax of causative constructions: cross-language similarities and divergences. In M. Shibatani, ed. *Syntax and semantics*, vol.6: 261-312. New York: Academic Press.

1981 *Language universals and linguistic typology*. Chicago: University of Chicago Press.

DE GUZMAN, Videa P.

1978 *Syntactic derivation of Tagalog verbs*. *Oceanic Linguistics* Special Publication No.16. Honolulu: The University Press of Hawaii.

1983 Ergative analysis for Philippine languages: an analysis. Paper read at the Third Eastern Conference on Austronesian Languages, Athens, Ohio. (To be published in *Studies in Austronesian Linguistics*).

FRANTZ, Donald

1981 Grammatical relations in universal grammar. Reproduced by the Indiana University Linguistics Club.

GERDTS, Donna

1983 Antipassives and causatives in Ilokano: evidence for an ergative analysis. Paper read at the Third Eastern Conference on Austronesian Languages, Athens, Ohio. (To be published in *Studies in Austronesian Linguistics.)*

GIBSON, Jeanne

1980 Clause union in Chamorro and in universal grammar. Ph.D. dissertation, University of California, San Diego.

McFARLAND, Curtis D.

1976 *A provisional classification of Tagalog verbs.* Institute for the Study of Languages and Cultures of Asia and Africa Monograph Series No.8. Tokyo Gaikokugo Daigaku.

O'GRADY, W.D.

1980 The universal characterization of passivization. *Linguistic Analysis* 6/4:393-405.

PERLMUTTER, David M., ed.

1983 *Studies in relational grammar 1.* Chicago: University of Chicago Press.

RAMOS, Teresita V.

1974 *The case system of Tagalog verbs. PL*, B-27.

SCHACHTER, Paul and Fe T. OTANES

1972 *Tagalog reference grammar.* Los Angeles: University of California Press.

De Guzman, V.P. "Some consequences of causative clause union in Tagalog". In Geraghty, P., Carrington, L. and Wurm, S.A. editors, *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics.* **C-93:59-72**. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.59*

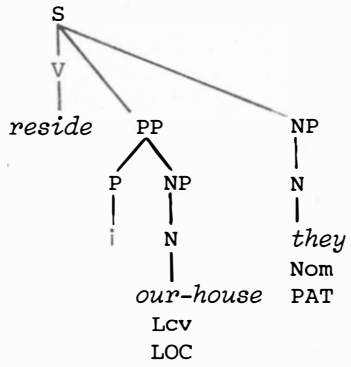
FOCUS AS RECENTRALISATION

Stanley Starosta

1. BACKGROUND[1]

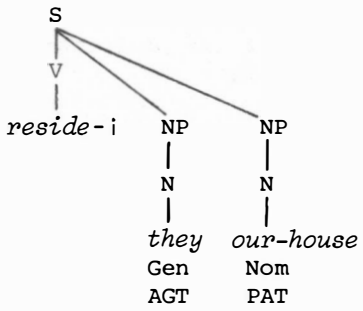
Three years ago, Starosta, Pawley, and Reid (hereafter SPR) presented a paper at the Bali conference in which they attempted to account for the evolution of Western Austronesian focus constructions as the result of a process of reinterpretation of nominalised equational constructions by analogy with functionally equivalent verbal constructions. For instance, assuming that PAN was an ergative language as defined in lexicase, so that the grammatical subject was always Patient, they posited intransitive constructions such as (1) whose verbs could undergo a lexical process of transitivisation which had the effect of reinterpreting the former Locus as a Patient, with a concomitant reinterpretation of the former Patient as Agent. The result would look like (2).

(1)



They reside in our house.

(2)



They inhabit our house.

In the lexicase framework in which the SPR paper was written, the Patient is obligatory in every clause (with certain statable exceptions), and constitutes the perceptual centre of each clause. The derivation process they posited can thus be considered 'recentralisation', the creation of a new verb whose perspective places the location of the action or state at the perceptual centre. The English glosses for (1) and (2) have been chosen to illustrate this difference in perspective. Thus *reside* is a verb which treats the residence as a Locus, but *inhabit* treats it as a Patient and the inhabitant as the grammatical Agent.

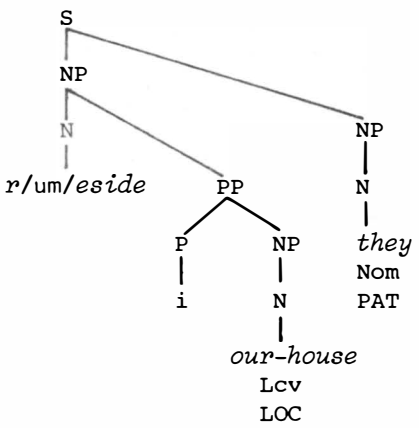
In addition to verbal constructions such as (1) and (2), SPR also posited lexical nominalisation rules in which the ancestors of the familiar 'focus affixes' such as -um- and -an functioned to nominalise verbs, as they still can

---

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 73-95. *Pacific Linguistics*, C-93, 1986.

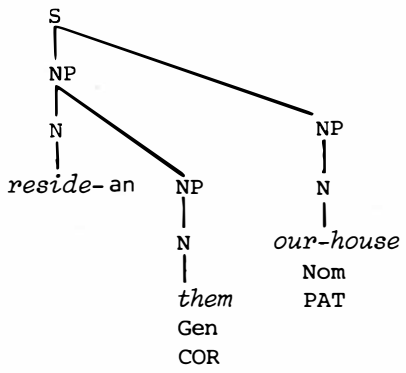
in modern Austronesian languages. These deverbal nouns could then function as predicates of equational sentences, resulting in examples such as (3) and (4):

(3)



They are the residents in our house.

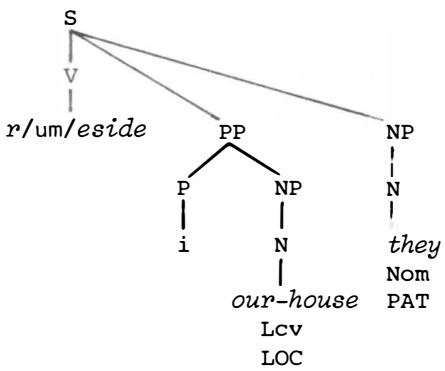
(4)



Our house is their residence.

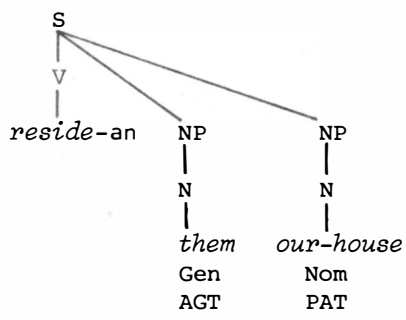
Note the close parallel in perspectives between (1) and (3) and between (2) and (4). In (1) and (3), the resident is the Patient, and in (2) and (4) the residence is the Patient. It was this syntactic and perceptual parallelism which SPR cited in support of the plausibility of their thesis that the nominal constructions (3) and (4) were reinterpreted by analogy with the verbal counterparts (1) and (2), resulting in (5) and (6), hybrid constructions with the syntactic analysis taken from the original verbal constructions and the morphology taken from the original deverbal nouns:

(5)



They reside in our house.

(6)

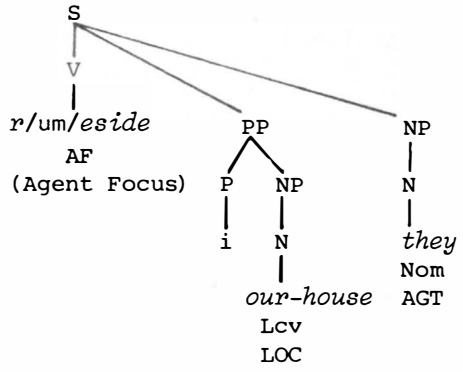


They inhabit our house.

Although SPR were trying to account for the evolution of focus in Philippine languages in particular, however, they did not quite get there. Their analysis resulted in a stage which was purely ergative, that is, one in which the grammatical Patient is always the grammatical subject. However, this is not the kind of system that has been hitherto assumed for Philippine languages analysed in case frameworks. In such analyses, it has normally been assumed that 'focus' is a kind of agreement affixation, so that -um- for example is a marker that signals the presence of the Agent case relation on the subject, as in (7), and -an frequently signals the presence of the Locus case relation on the subject, as in

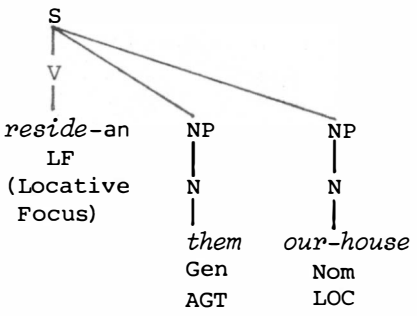
(8). Tagalog examples of this analysis based on Ramos' case grammar are given as (9) and (10):

(7)



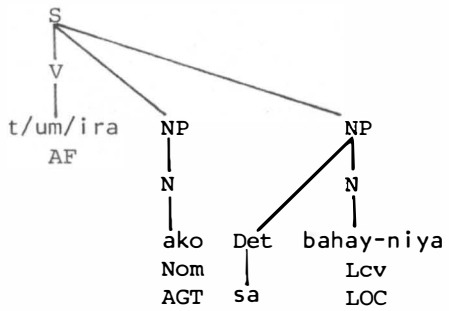
They reside in our house.

(8)



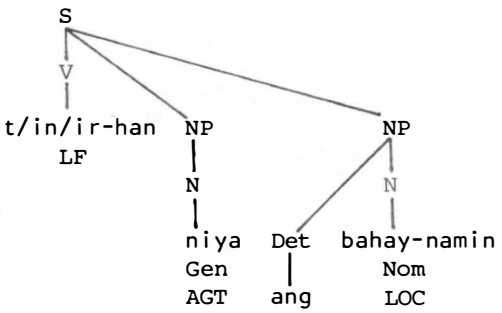
They inhabit our house.

(9) (cf. Ramos 1974:33)



I lived in his house.

(10) (cf. Ramos 1974:120)



He lived in our house.

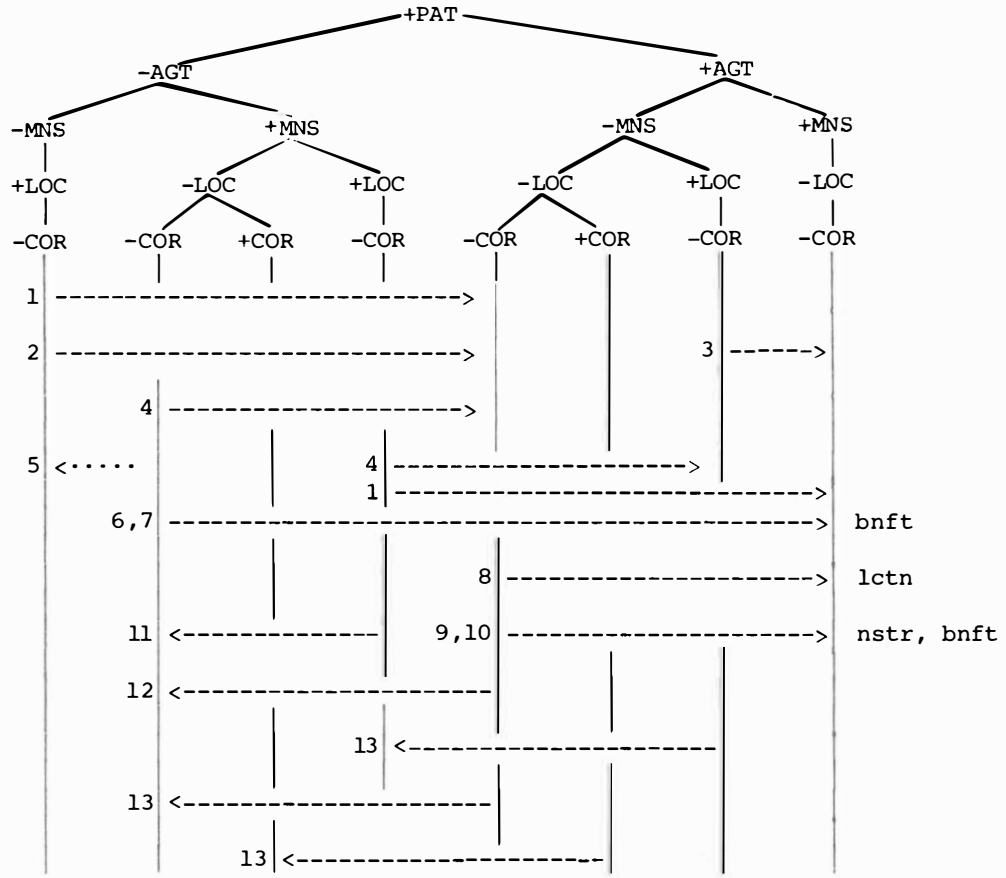
It would appear then that the SPR paper didn't really complete its work: it left us at the stage corresponding to (5) and (6), and didn't show us how to get across the last transition to (7) and (8). It will be my contention in this paper that that appearance is misleading, and that in fact SPR stopped exactly where they should have. That is, languages such as Tagalog are pure ergative languages, a conclusion that linguists such as Cena and De Guzman have also been approaching within the framework of relational grammar.

I will propose then that (5) and (6) represent the correct analysis of focus constructions, and that moreover focus is actually a process of lexical derivation that produces verbs with altered perspectives and/or different case frames. In particular, I will be concerned with a subset of these derivations in which there is a *recentralisation*, that is a change in perspective in which a new participant is chosen to be the perceptual centre of the sentence, that is, the Patient. From this point of view, a suffix such as -an signals to the hearer 1) that the verb has been derived into a new syntactic class, and 2) that the Patient subject of the sentence is to be interpreted selectionally as a location (at least for one set of verbs), just as a verb such as *died* signals to the hearer that the grammatical subject of a sentence such as *The flubadub died* is to be interpreted as animate.

In the next section of the paper I will be discussing the recentralisation rules I have posited so far for Tagalog, citing examples drawn mainly from De Guzman's *Syntactic derivation of Tagalog verbs* (hereafter TV for 'Tagalog verbs'). In the final section, I will offer morphological, syntactic, diachronic, and metatheoretical support for the recentralisation analysis.

2. ANALYSIS AND ILLUSTRATIONS

(11) Tagalog recentralisations



(11) is a tabulation of the recentralisation processes that I posit for Tagalog at this point. The tree at the top of the figure indicates the relevant syntactic classes I am assuming, defined in terms of case frame. Thus the +PAT at the top of the tree indicates that all the verb classes we are concerned with take a Patient in their case frames, as required by the lexicase Patient centrality hypothesis. The right branch of this tree covers all the verb classes which have an Agent in addition to a Patient in their case frames, etc.

The dashed arrows under the tree represent verbal derivation processes all of which involve recentralisation. For example, the arrow labelled (1) refers to a lexical process that applies to verbs in the intransitive locative class,

that is, verbs which have a Patient and an inner Locus in their case frames, and derives verbs which are simple transitives, that is, verbs with only Agent and Patient in their case frames. Exactly the same rule is shown further down in the diagram as deriving verbs from the intransitive locative means class to the transitive means class. All of the processes will now be illustrated.

(12) LDR-1

$$\begin{bmatrix} +V \\ -trns \\ +[+LOC] \\ \supset \begin{bmatrix} +LOC \\ \alpha F_i \end{bmatrix} \\ \supset \begin{bmatrix} +PAT \\ \beta F_j \end{bmatrix} \end{bmatrix} \quad >\text{---}> \quad \begin{bmatrix} +V \\ +trns \\ +lctn \\ \supset \begin{bmatrix} +PAT \\ \alpha F_i \end{bmatrix} \\ \supset \begin{bmatrix} +AGT \\ \beta F_j \end{bmatrix} \end{bmatrix}$$

$$\left\{ \begin{array}{lll} [C^1_o umX] & \rightarrow & [C^1_o Xan] \\ [mX] & \rightarrow & [Xan] \\ [maX] & \rightarrow & [maXan] \end{array} \right\}$$

Lexical derivation rule 1 is the one that would be needed to get from (5) to (6). It functions to reinterpret a Locus as a Patient, with a consequent reinterpretation of the former Patient as Agent. The suffix -an on the resulting verb signals the imposition of the selectional interpretation of 'location' on its new Patient. In the most recent version of the lexicase framework, the presence of an Agent is a necessary and sufficient condition for transitivity, so that LDR-1 always results in transitivisation. The operation of this rule is illustrated by the following examples. Numbers following the Tagalog examples refer to De Guzman 1978 unless otherwise noted. The analyses and sometimes the glosses are mine:

(13) LOC >---> PAT, PAT >---> AGT, lctn

a. nanggaling siya sa ibang bansa? [2:23]
1 PAT 3 LOC 5
2 4

He came from a foreign country.
PAT 1- -1 4 LOC
2 5

b. pinanggalingan na niya ang iba't ibang bansa? [2:23a]
1 2 AGT 4 PAT 6
3 5

He has visited other countries.
AGT 1- -1 5 LOC
3 6

c. bumili sila ng bahay sa 'Nu-West Realtors' [3:36]
1 PAT 3 MNS 5 LOC
2 4 6

They bought a house from Nu-West Realtors.
2 1- 3 4 -1 6

c'. *Jesse escaped from the Wells Fargo office with the money.*
PAT LOC MNS

d. binilhan ng nanay ng damit ang kaibigan [2:81]
1 2 AGT 4 MNS 6 PAT
3 5 7

Mother bought a dress from her friend.
3 1- 4 5 -1 7

d'. *Mother robbed her friend of a dress.*
AGT PAT MNS

The inner Locus of the input verb (ibang bansa *foreign country* in (13a)) is reinterpreted as the Patient (iba't ibang bansa *other countries* in (13b)). By my ergative analysis, this automatically results in the reappearance of the former grammatical Locus as the grammatical subject. The differences in the English glosses are intended to reflect the change of perspective that results from this derivation. Thus *foreign country* in the first gloss is a Locus, the source of the movement of the Patient *he*, but *other countries* in the second gloss is the Patient, that which is affected by the Agent *he*.

The same rule produces the transitive binilhan in (13d) from the intransitive bumili in (13c). Note that by my analysis, the notional 'object' ng bahay *a house* in (13c) is not a grammatical Patient, but rather an example of the Means case relation. This is consistent with relational grammar studies by Cena and De Guzman showing that such nominals are not treated grammatically as nuclear terms, that is, as Agent or Patient in the lexicase framework. Since English does not have good productive examples of this kind of pseudo-transitive construction (although certain European languages such as Finnish, Hungarian, and Russian apparently do), it is hard to convey the perspectual differences in the English gloss. Instead, I have provided the English examples (13c') and (13d') in an attempt to convey the flavour of the distinction that I think should be there.

(14) LDR-2

$$\begin{bmatrix} +V \\ -trns \\ +[+LOC] \\ \supset \begin{bmatrix} +PAT \\ \alpha F_i \end{bmatrix} \\ \supset \begin{bmatrix} +LOC \\ \beta F_j \end{bmatrix} \end{bmatrix} \quad >\text{---}> \quad \begin{bmatrix} +V \\ +trns \\ +afct \\ \supset \begin{bmatrix} +AGT \\ \alpha F_i \end{bmatrix} \\ \supset \begin{bmatrix} +PAT \\ \beta F_j \end{bmatrix} \end{bmatrix}$$

$$[C^1_o umX] \rightarrow [C^1_o Xin]$$

Lexical derivation rule 2 also functions to reinterpret a Locus of the verb (pumasok in (15a)) as a Patient with a consequent reinterpretation of the former Patient as Agent. The suffix -in (or its allomorph, the infix -in-) on the resulting verb signals the imposition of the selectional interpretation of 'affect' on its new Patient. This difference seems to be conveyed nicely by the gloss *drenched* for the derived verb:

(15) LOC >---> PAT, PAT >---> AGT, afct

a. pumasok ang tubig sa sapatos niya [4:66a]
1 2 PAT 4 LOC 6
3 5

The water went into his shoes.
2 PAT 1- -1 6 LOC
3 5

b. pinasok ng tubig ang sapatos niya [4:66b]
1 2 AGT 4 PAT 6
3 5

Water drenched his shoes.
AGT 1 6 PAT
3 5

(16) LDR-3

$$\begin{bmatrix} +V \\ +[+LOC] \\ \cup \begin{bmatrix} +PAT \\ \alpha F_i \end{bmatrix} \\ \cup \begin{bmatrix} +LOC \\ \beta F_j \end{bmatrix} \end{bmatrix} \quad >---> \quad \begin{bmatrix} +V \\ +dvrs \\ +[+MNS] \\ \cup \begin{bmatrix} +PAT \\ \beta F_j \end{bmatrix} \\ \cup \begin{bmatrix} +MNS \\ -dfnt \\ \alpha F_i \end{bmatrix} \end{bmatrix}$$

$$[maX] \rightarrow [maXan]$$

$$\left\{ \begin{matrix} [C_o^1 umX] \\ [iX] \\ [Xin] \end{matrix} \right\} \rightarrow [Xan]$$

Lexical derivation rule 3, like 1 and 2, also functions to reinterpret a Locus of the verb (pasyente in (17a)) as a Patient, but this time with a concomitant reinterpretation of the former Patient as Means. Since this rule does not add or remove an Agent from the case frame, it does not affect transitivity. Selectionally, it has the effect of imposing the interpretation of 'location' on the new Patient. In (17a), that is, the tumour is interpreted as the perceptual centre, but in (17b), it is the patient who is (fortuitously) the grammatical Patient, the percentual centre selectionally interpreted as a location. This interpretation is again signalled by the suffix -an on the verb nakitaan. The posited perspectual difference is reflected in the English example (17b'), where *amphetamines* might be regarded as the notional 'object', although grammatically the direct object is *the nurse*:

(17) LOC >---> PAT, PAT >---> MNS, lctn

a. nakita ng doktor sa pasyente ang isang malaking tumor [3:24]
1 2 AGT 4 LOC 6 PAT 8 9
3 5 7

The doctor saw a big tumour in the patient.
2 AGT 1 7 8 PAT 4- -4 LOC
3 9 5

b. nakita?an ng doktor ang pasyente ng isang malaking tumor [3:24a]
1 2 AGT 4 PAT 6 MNS 8 9
3 5 7

The doctor saw the patient as having a big tumour.
2 3 1 4 5 7 8 9

b'. *The doctor caught the nurse with some stolen amphetamines.*
AGT PAT MNS

(18) LDR-4

$$\begin{bmatrix} +V \\ -trns \\ +[+MNS] \\ \supset \begin{bmatrix} +PAT \\ \alpha F_i \end{bmatrix} \\ \supset \begin{bmatrix} +MNS \\ -dfnt \\ \beta F_j \end{bmatrix} \end{bmatrix} \quad >---> \quad \begin{bmatrix} +V \\ +trns \\ +afct \\ \supset \begin{bmatrix} +AGT \\ \alpha F_i \end{bmatrix} \\ \supset \begin{bmatrix} +PAT \\ \beta F_j \end{bmatrix} \end{bmatrix}$$

$[C^1_o umX] \rightarrow [C^1_o Xin]$

Lexical derivation rule 4 functions to upgrade the inner Means of the verb (gumawa in (19a)) to a Patient, with a consequent reinterpretation of the former Patient as Agent. As indicated by the gloss, the main function of this rule is to make it possible to mark the former notional 'object' as definite, which is not possible with an inner Means actant. The suffix -in on the resulting verb here primarily signals transitivity, which semantically adduces the interpretation of an external Agent acting directly on a definite Patient.

(19) MNS >---> PAT, PAT >---> AGT, afct

a. gumawa? si Angel ng tugtugin [2:12]
1 2 PAT 4 MNS
3 5

Angel composed a piece of music.
PAT 1 4 MNS
3 5

b. gagawa?in ni Angel ang tugtugin [cf. 4:73a]
1 2 AGT 4 PAT
3 5

Angel will compose the piece of music.
AGT 1- -1 4 PAT
3 5

c. humiram ang estudyante ng libro sa guro? [2:8a]
1 2 PAT 4 MNS 6 LOC
3 5 7

The student borrowed a book from the teacher.
2 3 1- 4 5 -1 6 7

c'. *The student absconded from the library with a book.*
PAT LOC MNS

d. hiniram ng estudyante ang libro sa guro? [2:8b]
1 2 AGT 4 PAT 6 LOC
3 5 7

The student borrowed the book from the teacher.
2 AGT 1- 4 PAT -1 6 LOC
3 5 7

(20) LDR-5

$$\begin{bmatrix} +V \\ +[+MNS] \\ \supset \begin{bmatrix} +MNS \\ -dfnt \\ \alpha F_i \end{bmatrix} \end{bmatrix} \text{ >---> } \begin{bmatrix} +V \\ +[+LOC] \\ \supset \begin{bmatrix} +LOC \\ +dfnt \\ \alpha F_i \end{bmatrix} \end{bmatrix}$$

LDR-5 differs from the previous four derivation rules in that it does not involve reinterpretation of the Patient. Instead, it reinterprets an inner Means, which is required to be indefinite, as inner Locus, which with this class of verbs is required to be definite. This enables notional 'objects' to be interpreted as definite without being subjects. It thus allows personal pronouns and proper nouns, which are lexically definite, to appear as non-nominative 'objects'. This property is highly reminiscent of the use of a locative case form to mark definite direct objects in Spanish, but there is no reason to assume there has been Spanish influence here, since the same thing occurs in Hindi and in Amis (Chen 1985), a Formosan language with no history of contact with Spanish. Since no morphophonemic modifications are involved, this rule is an example of zero derivation.

(21) MNS >---> LOC, dfnt

a. maghihintay siya ng magpeperyodiko sa kanto [2:7a]
1 PAT 3 MNS 5 LOC_o
2 4 6

He will wait for a newsboy at the corner.
PAT 1- -1- -1 3 MNS 5- -5 LOC_o
2 4 6

b. maghihintay siya sa magpeperyodiko sa kanto [2:7b]
1 PAT 3 LOC_i 5 LOC_o
2 4 6

He will wait for the newsboy at the corner.
PAT 1- -1- -1 3 LOC_i 5- -5 LOC_o
2 4 6

The LOC_i and LOC_o refer to inner and outer case relations respectively.

(22) LDR-6

$$\begin{bmatrix} +V \\ -trns \\ -[+LOC] \\ +[+MNS] \\ \cup \begin{bmatrix} +PAT \\ \alpha F_i \end{bmatrix} \end{bmatrix} \quad >---> \quad \begin{bmatrix} +V \\ +trns \\ +bnft \\ -[+LOC] \\ +[+MNS] \\ \cup \begin{bmatrix} +AGT \\ \alpha F_i \end{bmatrix} \end{bmatrix}$$

$$[C^1_o umX] \rightarrow [C^1_o Xan]$$

LDR-6 in effect captures an outer Correspondent case relation, reinterprets it as Patient, and assigns it a beneficiary interpretation. The new Patient crowds out the old one, which is reinterpreted as Agent, thereby resulting in transitivisation. The idea of 'capture' however is not directly reflected in this formalisation, since outer case relations can't be referred to in the case frame. Thus it can equally be seen as a rule which creates a new Patient and interprets it as a beneficiary, while reinterpreting the old one as the Agent.

(23) Ø >---> PAT, PAT >---> AGT, bnft

a. bumili ako ng silya para sa nanay [2:52]
1 PAT 3 MNS 5 6 COR_o
2 4 7

I bought a chair for mother.
2 1 3 4 5 7

a'. *Oliver absconded with a chair for Fagan.*
PAT MNS COR_o

b. binilhan ng nanay ng damit ang kaibigan [2:81]
1 2 AGT 4 MNS 6 PAT
3 5 7

Mother bought a dress for her friend.
3 1- 4 5 -1 7

b'. *Mother presented her friend with a dress.*
AGT PAT MNS

(24) LDR-7

$$\begin{bmatrix} +V \\ -trns \\ \cup \begin{bmatrix} +PAT \\ \alpha F_i \end{bmatrix} \end{bmatrix} \quad >---> \quad \begin{bmatrix} +V \\ +trns \\ +bnft \\ \cup \begin{bmatrix} +AGT \\ \alpha F_i \end{bmatrix} \end{bmatrix}$$

$$[C^1_o umX] \rightarrow [iC^1_o X]$$

$$[mpagX] \rightarrow [ipagX]$$

LDR-7 is very similar in function to LDR-6, except that it is not restricted to non-locus means intransitives. It also differs in its morphological effect, since it assigns an i- prefix rather than an -an suffix. I follow De Guzman in treating the prefix mag- as morphophonemically composed of pag- preceded by m-, since this simplifies the statement of this and other rules and captures the relation between mag- and -um- more directly.

(25) Ø >---> PAT, PAT >---> AGT, bnft

a. nagluto? siya ng litson [cf. 2:28]
1 PAT 3 MNS
2 4

He cooked a roast.
2 1 3 4

a'. *They built with lannon stone.*
PAT MNS

b. ipinagluto ko ng litson ang nanay [cf. 2:53]
1 AGT 3 MNS 5 PAT
2 4 6

I cooked mother a roast.
AGT 1 PAT 3 MNS
2 6 4

(26) LDR-8

$$\begin{bmatrix} +V \\ +trns \\ +afct \\ \supset \begin{bmatrix} +PAT \\ \alpha F_i \end{bmatrix} \end{bmatrix} >\text{---}> \begin{bmatrix} +V \\ +trns \\ +lctn \\ +[+MNS] \\ \supset \begin{bmatrix} +MNS \\ -dfnt \\ \alpha F_i \end{bmatrix} \end{bmatrix}$$

in] → an]

LDR-8 in effect captures the outer Locus of a transitive affect verb, reinterpreting it as a locational Patient. Since the verb is already transitive, the displaced former Patient is reinterpreted as Means. As in the case of other rules in which a former Patient is downgraded to Means, the new Means actant is interpreted as indefinite. This rule applies only to -in-suffixed stems, and replaces the -in by -an.

(27) Ø >---> PAT, PAT >---> MNS, lctn

a. kaka?inin ni Melissa ang mansanas [4:73a]
1 2 AGT 4 PAT
3 5

Melissa will eat the apple.
AGT 1- -1 4 PAT
3 5

b. kina?inan nila ng hapunan ang bagong restawrang ito kagabi [2:26a]
1 AGT 3 MNS 5 PAT 7 8 LOC_O
2 4 6 9

This new restaurant was eaten dinner at by them last night.
8 6 7 1- -1- 4 -1 2- -2 9

b'. *The IRS deprived Chrysler of its unearned profits last year.*
AGT PAT MNS LOC_O

(28) LDR-9

$$\begin{bmatrix} +V \\ +trns \\ +afct \\ \supset \begin{bmatrix} +PAT \\ \alpha F_i \end{bmatrix} \end{bmatrix} \text{ >---> } \begin{bmatrix} +V \\ +trns \\ +nstr \\ +[+MNS] \\ \supset \begin{bmatrix} +MNS \\ -dfnt \\ \alpha F_i \end{bmatrix} \end{bmatrix}$$

[Xin] → [iX]

LDR-9 is quite similar to LDR-8, except that LDR-9 captures an outer Means case relation (an instrument) and interprets its new Patients as instrumental rather than locational. Again, the former Patient must be reinterpreted to appear as an indefinite inner Means 'object'. The morphophonemic effect is to remove the -in suffix and add an i- prefix.

(29) Ø >---> PAT, PAT >---> MNS, bnft

a. pinutol ni Pablo ang damo ng karit [2:47b]
1 2 AGT 4 PAT 6 MNS_O
3 5 7

Pablo cut the grass with a sickle.
AGT 1 4 PAT 6 MNS_O
3 5 7

b. ipuputol niya ito ng tubo [3:1b]
1 AGT PAT 4 MNS
2 3 5

He will cut this person some sugarcane.
2 1- -1 3 4 5

b'. *She availed herself of a handy microcomputer.*
AGT PAT MNS

(30) LDR-10

$$\begin{bmatrix} +V \\ +trns \\ -[+LOC] \\ \supset \begin{bmatrix} +PAT \\ \alpha F_i \end{bmatrix} \end{bmatrix} \quad >\text{---}> \quad \begin{bmatrix} +V \\ +trns \\ +bnft \\ -[+LOC] \\ +[+MNS] \\ \supset \begin{bmatrix} +MNS \\ -dfnt \\ \alpha F_i \end{bmatrix} \end{bmatrix}$$

[Xin] → [iX]

LDR-10 is again similar in function to LDR-8, except that LDR-10 applies to non-locational transitive verbs and 'centralises' outer Correspondent actants, reinterpreting them as beneficiaries. Again, the former Patient is crowded out by the new one, and is reinterpreted as an indefinite inner Means. This rule applies only to -in-suffixed stems, and removes the -in suffix while adding an i- prefix.

(31) Ø >---> PAT, PAT >---> MNS, nstr

a. pinutol ni Pablo ang damo ng karit [2:47b]
1 2 AGT 4 PAT 6 MNS_O
3 5 7

Pablo cut the grass with a sickle.
AGT 1 4 PAT 6 MNS_O
3 5 7

b. inpuputol niya ito ng tubo [3:1a]
1 AGT PAT 4 MNS
2 3 5

He will cut sugarcane with this.
2 1- -1- 5 -1 3

b'. *She availed herself of a handy microcomputer.*
AGT PAT MNS

(32) LDR-11

$$\begin{bmatrix} +V \\ -trns \\ +[+LOC] \\ \supset \begin{bmatrix} +PAT \\ \alpha F_i \end{bmatrix} \\ \supset \begin{bmatrix} +LOC \\ \alpha F_i \end{bmatrix} \end{bmatrix} \quad >\text{---}> \quad \begin{bmatrix} +V \\ -trns \\ \supset \begin{bmatrix} +PAT \\ +plrl \\ \alpha F_i \end{bmatrix} \end{bmatrix}$$

[mpagX] → [pagXan]

LDR-11 derives reciprocal verbs from directional intransitives. It reinterprets the Locus actant of the input verb as a Patient by in effect incorporating it into the Patient referent of the new verb. The rule expresses this by applying only to locational intransitives which impose the same set of selectional restrictions on Locus and Patient, and by deriving a verb which takes no inner Locus (shown by the absence of the +[+LOC] feature on the right side of the arrow) and interprets its new Patient as plural. Morphophonemically, the rule removes the initial m- prefix from the pag stem and adds an -an suffix.

(33) LOC >---> PAT, rcpr

a. magbibigay ang propesor ng selyo sa mga estudyante
1 2 PAT 4 MNS 6 7 LOC
3 5 8

The professor will give of (his) stamps to the students.
2 PAT 1- -1 4 MNS 6- -6 7-8
3 5

b. pagbibigyan ang mga estudyante ng selyo
1 2 3 PAT 5 MNS
4 6

The students will present each other with stamps.
2 PAT 1- -1- -1- -1 5 MNS
3-4 6

(34) LDR-12

$$\begin{bmatrix} +V \\ +trns \\ \supset \begin{bmatrix} +PAT \\ \alpha F_i \end{bmatrix} \\ \supset \begin{bmatrix} +AGT \\ \beta F_j \end{bmatrix} \end{bmatrix} \text{ >---> } \begin{bmatrix} +V \\ -trns \\ \supset \begin{bmatrix} +MNS \\ -dfnt \\ \alpha F_i \end{bmatrix} \\ \supset \begin{bmatrix} +PAT \\ \beta F_j \end{bmatrix} \end{bmatrix}$$

$$\left\{ \begin{bmatrix} iX \end{bmatrix} \\ \begin{bmatrix} Xan \end{bmatrix} \\ \begin{bmatrix} Xin \end{bmatrix} \right\} \rightarrow [mpagX]$$

LDR-12 is a derivation rule which has the effect of deriving an intransitive verb from a transitive one. In the process, the former Patient is 'demoted' into an indefinite inner Means relation, and the former Agent takes over the Patient role. The result is thus an example of what is traditionally referred to as an 'anti-passive' construction, although here this is not an isolated process, but one of a number of processes which displace a Patient and reinterpret it as an indefinite inner Means 'object'. The rule applies to i-, -an, or -in stems, and replaces the prefix or suffix by a mpag- prefix.

(35) AGT >---> PAT, PAT >---> MNS, anti-passive

a. inilatag niya ang banig sa araw [3:48]
```
   1     AGT  3   PAT  5  LOC
          2        4       6
```
He spread out the mat in the sun.
```
AGT  1-   -1  3  PAT 5-  -5  LOC
 2                 4          6
```

b. naglatag siya ng banig sa araw [3:49]
```
   1     PAT  3   MNS  5  LOC
          2        4       6
```
He spread out a mat in the sun.
```
 2   1-   -1  3  4  5-  -5   6
```

b'. *He covered up in the arctic with heavy blankets.*
```
PAT                    LOC                  MNS
```

c. patatakbuhin ko ang bata? [De Guzman MSa:30]
```
     1       AGT  3   PAT
              2        4
```
I will race the child.
```
AGT  1-   -1   3    4
 2
```

d. magpapatakbo ako ng bata? [De Guzman MSa:30]
```
     1       PAT  3   MNS
              2        4
```
I will race with a child.
```
PAT  1-   -1   3      MNS
 2                     4
```

e. ipinaluto? ko sa nanay ang manok [De Guzman MSa:29]
```
     1     AGT  3   COR   5   PAT
            2        4         6
```
I had mother cook the chicken.
```
2  1-    4     -1   5     6
```

e'. *I compared the chicken to mother.*
```
AGT                 PAT        COR
```

f. nagpaluto? ako sa nanay ng manok
```
    1      PAT  3   COR   5   MNS
            2        4         6
```
I had mother cook (a) chicken.
```
2  1-    4     -1   5     6
```

f'. *I imposed on the secretary with my frequent rush requests.*
```
AGT                    COR                                MNS
```

(36) LDR-13

$$\begin{bmatrix} +V \\ +psyc \\ +trns \\ \supset \begin{bmatrix} +PAT \\ \alpha F_i \end{bmatrix} \\ \supset \begin{bmatrix} +AGT \\ \beta F_j \end{bmatrix} \end{bmatrix} \quad >---> \quad \begin{bmatrix} +V \\ +psyc \\ -trns \\ +[+MNS] \\ \supset \begin{bmatrix} +PAT \\ \beta F_j \end{bmatrix} \\ \supset \begin{bmatrix} +MNS \\ -dfnt \\ \alpha F_i \end{bmatrix} \end{bmatrix}$$

[ma → [maka

LDR-13 is the anti-passive rule which applies to psychological ma- stems, replacing the ma- by maka-. As in the case of the previous rule, the former Patient is downgraded to an indefinite Means 'object', and the former Agent is reinterpreted as the necessarily definite Patient.

(37) AGT >---> PAT, PAT >---> MNS, anti-passive

a. nakita ng mga bata? ang salamangkero [3:98]
1 2 AGT 5 PAT
3 6

The children saw the magician.
2 AGT 1 5 PAT
3 6

b. nakakita ang mga bata? ng salamangkero [3:99]
1 2 PAT 4 MNS
3 5

The children saw a magician.
2 3 1 4 5

b'. *The transformationalists tasted of the Pierian waters.*
PAT MNS

3. JUSTIFICATION

3.1 Comparative and historical justification

3.1.1 Simplified account of the transition from PAN to Philippine focus systems

(a) If Tagalog is a pure ergative language, then we can eliminate one step in the SPR account of the evolution of focus in Austronesian.

(b) SPR reconstructed PAN as a mixed accusative-ergative language rather than as a pure ergative system mainly by analogy with De Guzman's 'TV' analysis of Tagalog. However, they did not specifically reconstruct any accusative constructions for the proto-language. If it turns out that Tagalog can in fact now be analysed as a pure ergative language, it is quite possible that the same analysis can be

applied to PAN resulting in a simpler proto-system and a simpler account of the transition from PAN to Philippine systems.

(c) The recentralisation analysis of focus proposed here applies without any alterations to accusative Austronesian languages such as Kara as well as to ergative languages like Tagalog. This means that only a single simple transition need be posited in the evolution of such languages: a shift from a centrality-based subject-choice hierarchy to a salience-based one. A plausible mechanism for such a change in Oceanic languages was in fact outlined at the end of the SPR paper.

(d) The Chamorro connection

(i) The ergative/recentralisation analysis proposed in this paper reveals a clear syntactic and morphological distinction between transitive and intransitive verb classes which did not show up in earlier case grammar analyses such as TV. This distinction seems to be the same one that is found much more strikingly in Chamorro, and may have implications for establishing the position of Chamorro in the Austronesian family tree and reconstructing the grammatical system of the common ancestor.

(ii) This transitivity distinction in turn was important in the process of 'aux axing' that SPR posited to account for the evolution of Oceanic accusative systems from the posited ergative system. Chamorro has also undergone a similar process if not the same one. The diachronic implications of this fact seem paradoxical, but at least the present analysis has helped to point out the problem.

(iii) This analysis has also established a link in Tagalog between morphological plural marking and the class of syntactically intransitive verbs, a phenomenon which again has a close parallel in Chamorro. Again, the diachronic implications remain to be investigated, but the phenomenon has at least been identified for further study.

3.2 Metatheoretical justification

3.2.1 Simplification and elimination of adhociery

The TV analysis uses features such as [±erg] to account for differences of derivational potential. These are in effect rule features, and thus constitute an admission of a failure to find a motivated explanation. In the present analysis, such features are replaced by the transitivity feature, which is amply motivated in terms of syntactic and morphological criteria.

3.2.2 Universal statements

3.2.2.1 Ergativity

(a) The analysis proposed here is consistent with, and thus indirectly lends support to, the standard lexicase universal definition of an ergative language as one in which Patient takes priority over Agent in subject choice.

(b) The ergative/recentralisation analysis stated here is compatible with, and lends further support to, the lexicase 陰 陽 [yīn-yáng] theory of ergative-accusative typology, i.e. that languages choose to mark their subjects either on

the basis of saliency (AGT takes precedence over PAT, i.e. 陽) or centrality (PAT takes precedence over AGT, i.e. 陰).

(c) This analysis lends indirect support to the lexicase analysis of intransitive sentences in accusative languages such as English as having Patient subjects. The reasoning is that this analysis captures the fact that intransitive sentences in both ergative and accusative languages are grammatically identical.

(d) The lexicase analysis of ergativity automatically explains why (minimally marked) causative verbs in Philippine and Formosan languages are generally 'object focus': causativisation necessarily results in a transitive verb, and the subject of a transitive verb in an ergative construction is Patient.

3.2.2.2 Case

(a) The inventory of five case relations employed in this analysis is (except for TIM) the same set proposed as universals in Starosta 1982 primarily on the basis of evidence from English, and the definitions and uses of the CR's are similar. Thus this analysis lends support to the claims of universality for this list of primitives.

(b) Lexicase requires the ability to refer to a category of 'actor' defined as 'Agent if present, otherwise Patient'. The drawback of this definition is that it is a disjunction, but the advantage is that exactly the same definition works for ergative and accusative languages.

(c) The TV analysis, which proposes a mixed ergative-accusative analysis, is replaced here by a pure ergative system, which is conceptually simpler, and suggests the possibility that all reputedly mixed systems can be so reanalysed.

3.2.3 Patient centrality

(a) The present analysis is consistent with a universal hypothesis that Patient centrality applies in all human languages.

(b) In De Guzman's account of Raising in Tagalog, an argument is in effect raised into an empty subject position in a transitive clause. In a lexicase account based on the present ergative analysis, the argument is 'raised' to PAT rather than subject. Stated in this way, the same principle applies equally to accusative languages such as English, and in fact in English at least it applies to transitive as well as intransitive clauses.

(c) Complementation

(i) Given the present Patient Centrality analysis, a general statement can be made that sentential complements substitute for PAT arguments in transitive clauses, in ergative as well as accusative languages.

(ii) Similarly, a general statement can be made that direct and indirect quotes substitute for PAT arguments in transitive and perhaps also intransitive clauses in both ergative and accusative languages.

3.2.4 Lexicase versus Relational Grammar

3.2.4.1 Power

(a) De Guzman's RG analysis of anti-passivisation must refer to 'final' terms, and while this seems preferable to an analysis which refers to more than one stratum, it is still global reference. A lexicase analysis has only one 'stratum', and so such global reference is excluded in principle. The lexicase analysis is thus less powerful, and so metatheoretically preferable.

(b) After proving that RG metatheory must allow for Patient subjects in addition to the usual Agent subjects in order to account for Tagalog syntax, Cena concludes that probably the metatheory should allow for a given language to have its subjects take other semantic roles as well. However, the lexicase system is more constrained in allowing only two possible subject choices, Patient or Agent, rather than an open list as Cena suggested.

3.2.4.2 Inventory of primitive elements

(a) RG analyses of ergative syntax must posit the ad hoc Grammatical Relations 'Ergative' and 'Absolute' or their equivalents in order to capture certain important generalisations. In the lexicase analysis, however, no special new equipment is needed to account for ergative systems; instead, exactly the same inventory of case relations and case forms can be employed in both kinds of systems. In particular, 'subject' has the same role in both kinds of typology, whereas in De Guzman's analysis, Subject is specialised to accusative systems.

(b) Unlike De Guzman's RG analysis, the lexicase approach proposed here makes no grammatical distinction between formally identical ergative and genitive case forms. The necessity for making such a distinction in the first place seems to be again the result of RG's failure to distinguish between case forms and case relations.

3.2.4.3 Naturalness

Both Cena's and De Guzman's RG analyses run up against the need for the ability to refer to semantic roles in order to express important grammatical generalisations. However, Relational Grammar metatheory does not provide such a capacity, and thus is shown to be inadequate as a theory of natural language, at least as presently constituted. From the lexicase point of view, of course, this is due to the RG confusion of case forms and case relations. If RG is modified in accordance with De Guzman's suggestions to allow for direct reference to semantic roles, what will have been created is a variant of lexicase.

3.2.4.4 Generalisation

(a) Intransitives have the same representation in ergative and accusative languages.

(b) De Guzman notes that her RG anti-passive analysis closely parallels the RG passive analysis for accusative languages. Lexicase goes even further, however, in analysing the outputs of both processes to be not just parallel but grammatically identical: intransitive clauses with Patient subject and Means adjuncts.

(c) The RG anti-passive rule applies only in ergative languages. However, the corresponding lexicase rule, stated in terms of case relations rather than 'grammatical relations', applies without any modification to accusative languages as well. An English example is the derivation of the intransitive verb *chew on* from the transitive verb *chew*.

(d) The lexicase analysis accounts for the chômeurique properties of non-subject 'direct objects' in Tagalog by analysing them as having a non-nuclear case relation, Means. This analysis is superior to the RG analysis because it motivates the choice of case form: Means occurs in the Genitive case form in Tagalog. A weak point of a chômeur analysis is that chômeurs by definition have no grammatical relation, so there is no reason for them to occur in one case form rather than some other one.

(e) De Guzman's Relational Grammar analysis has no explanation of why non-subject Agents should act as 'nuclear terms', primarily, I think, because the status of case relations such as Agent and Patient in RG is not clear (as De Guzman herself points out). In lexicase this fact is an automatic consequence of a universal definition of 'nuclear term': Agent and Patient are the 'nuclear terms' of lexicase grammars regardless of whether they are subject or object, that is, regardless of whether the syntactic type is ergative or accusative.

(f) The recentralisation analysis posits many processes of the schematic form $PAT \rightarrow CR_i$, $CR_j \rightarrow PAT$ which operate identically in transitive and intransitive clauses. In a relational grammar analysis in which the sole argument of an intransitive clause is a 1, I think this is not possible, though it may work in De Guzman's 2-centred approach.

3.3 Morphological justification

(a) As in De Guzman's RG analysis, the lexicase Patient Centrality analysis reveals a clear ±transitive distinction that matches very nicely with categories of affixation.

(b) Similarly, both the lexicase and RG analyses show the result of antipassivisation to be an intransitive clause, which again matches the morphological evidence.

(c) The lexicase approach is able to unify at least some of the functions of 'focus affixes' which were treated as separate and unrelated in TV. Thus the verbal prefix i- in the present analysis has only one function: to signal the reinterpretation of an outer non-Locus as Patient, whereas in TV and in other case-type analyses, including tagmemic treatments, the functions had to be listed as separate and unrelated.

3.4 Semantic justification

(a) The lexicase analysis makes the claim that the same situation or action may be conceptualised in different ways by speakers of different languages depending on the range of verbal constructions available to them. Cena on the other hand proceeds from the assumption that conceptualisation is identical for speakers of all languages. This question cannot be settled in this paper, but at least there is a clear difference in empirical claims that is at least in principle subject to empirical resolution.

(b) Lexicase appears to provide a more principled account of semantic differences in different focusses. For example, De Guzman cites the following example:

Tinirahan nila ang bahay namin

glossed as *Our house was lived in by them.* In this LF construction, she says, the attention is on what action was done to the house or what ensues from it. This meaning increment can easily be connected with the lexicase analysis of LF constructions in which the house is a Patient being acted on by an Agent.

(c) When a given verb in a particular focus may occur in more than one shape, e.g. the two OF forms of *cook*, lulutuin and ilulutu, the lexicase analysis predicts a particular perspectual difference between them. Once again, I have no evidence bearing on this question as yet, but at least there is again a clear difference in empirical claims that is in principle subject to empirical resolution.

3.5 Syntactic justification

3.5.1 Simplicity

(a) The present lexicase analysis posits a smaller inventory of case relations and a simpler system CF-CR mapping than TV or other case analyses did.

(b) The Subject choice rule in particular is extremely simple: instead of the usual 'focus' case analysis in which the subject may bear any of a large number of case relations, this analysis allows it to bear only one: Patient.

3.5.2 Generality

(a) The pure ergative analysis posited for Tagalog is conceptually simpler than the mixed accusative-ergative one proposed in TV, and thus better if it really works.

(b) In TV, several derivational rules refer to the same feature disjunction, [-erg] and [-pot, +act]. These are ad hoc rule features, and the generalisation they are missing is captured in the present analysis by replacing them with the well-motivated [-transitive] feature.

3.5.3 Inventory

(a) The present inventory replaces the rather large TV case relation inventory by a very short one, with different verbs imposing different selectional implications on the same limited set of CRs.

3.5.4 Derivation

(a) By treating 'focus' as derivation rather than as inflection, the present analysis is able to explain certain forms in which inflectional affixes appeared to be inside of derivational affixes in the TV analysis, in contradiction to usual morphological criteria for distinguishing the two types of word formation.

4. CONCLUSION

The Relational Grammar and lexicase ergative analyses are superior in a number of respects to previous Fillmorean case grammar analyses. It is interesting to note that RG as practised by Cena and De Guzman seems to be converging with lexicase. This is a result of greater attention to semantic roles on the part of the former investigators and the semantic bleaching of case relations in lexicase. In terms of capturing language-specific and universal generalisations, however, lexicase seems to be (at least temporarily) in the lead.

NOTE

1. I would like to thank Josie Clausen, Vida De Guzman, and Terry Ramos for reading and commenting on earlier versions of this paper.

REFERENCES

CENA, R.M.

1977 Patient primacy in Tagalog. Annual meeting, Linguistic Society of America, Los Angeles, 28-30 December.

CHEN, Teresa M.

1985 *Verbal constructions and verbal classification in Nataoran-Amis*. (Ph.D. dissertation, 1982, University of Hawaii.) *PL*, C-85.

De GUZMAN, Videa P.

1978 *Syntactic derivation of Tagalog verbs*. *Oceanic Linguistics* Special Publication No.16. Honolulu: University Press of Hawaii.

1979 Morphological evidence for primacy of patient as subject in Tagalog. Annual meeting, Linguistic Society of America, Los Angeles, 27-29 December.

1983 Ergative analysis for Philippine languages: an analysis. Paper presented at the Third Eastern Conference on Austronesian Languages, Athens, Ohio.

1986 Some consequences of causative clause union in Tagalog. In this volume, 59-72.

GIBSON, Jeanne D.

1980 Clause union in Chamorro and universal grammar. Ph.D. dissertation, University of California, San Diego.

RAMOS, Teresita V.

1974 *The case system of Tagalog verbs*. *PL*, B-27.

STAROSTA, Stanley

1982 Case relations, perspective, and Patient centrality. *University of Hawaii Working Papers in Linguistics* 14/1:1-34.

STAROSTA, Stanley, Andrew K. PAWLEY and Lawrence A. REID

1982 The evolution of focus in Austronesian. In S.A. Wurm and Lois Carrington, eds *Papers from the Third International Conference on Austronesian linguistics*, vol.2: *Tracking the travellers*, 145-170. *PL*, C-75.

n.d. The evolution of focus in Austronesian. Department of Linguistics, University of Hawaii. MS. Expanded version of Starosta, Pawley and Reid 1982.

WORD ORDER CHANGE IN MALAY

Susanna Cumming

1. INTRODUCTION

This paper will concern the change from the predominantly predicate-initial[1] word order of Proto-Austronesian to the predominantly argument-initial order of modern Malay/Indonesian.[2] I will first consider in some detail the more accessible stage of this change, by discussing the differences in the word order characteristics of a text written in Early Modern Malay (circa 1900), and a text in informal spoken Modern Indonesian (circa 1982); subsequently, I will venture some hypotheses on how the Early Modern Malay stage might have emerged through a process of functional reanalysis from an earlier predicate-initial stage.

While both my older Malay and my modern Indonesian texts are predominantly argument-first, in the Malay text the predicate precedes its arguments 25% of the time, while in the Indonesian text this is true only 12% of the time. This fact reflects a difference in the pragmatic value of prepredicate versus post-predicate position for NPs. Since we know that Malay came originally from a predicate-initial language, this small statistical change can be seen as merely a continuation of a change that has been going on for hundreds of years and has already resulted in a major word order shift. I will suggest that for the Malay and Indonesian data I looked at, the most natural explanation for the shift is in terms of a growing tendency (favoured by a number of morphological changes) for agents to be prepredicate, and for a corresponding restriction in the use of the "patient-trigger" construction.

1.1 Data

As mentioned above, the data for this study consist of narrative texts. The texts were divided into clauses, and for each clause information was recorded (using a computer database system) about the word order of the clause and how it was connected to other clauses, the verb morphology, and the semantic role, type, and number of clauses since last mention ("lookback") of each of the arguments.

My source for Malay is an Early Modern Malay text entitled *Tarikh Datuk Bentara Luar Johor*, which has been transliterated, edited, annotated, and translated by Amin Sweeney (1980a and b). This is a collection of autobiographical writings by Mohamed Salleh bin Perang (1841-1915), a Malay statesman. While there are earlier prose writings in Malay, this work is notable for being (according to Sweeney) the earliest by a native speaker of Malay which is aimed at

---

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 97-111. *Pacific Linguistics*, C-93, 1986.

a Malay, rather than a European, audience, and which is written to be read individually rather than chanted (in a "rhythmic monotone" - Sweeney 1980a) to an audience. I coded 337 clauses from two different parts of this source; 106 were from an official account of Salleh's life written in a relatively formal style, and 231 from an account in a letter to a friend of a war Salleh fought in.

My Indonesian data are taken from two oral narratives produced by Wilson Manik and collected and transcribed by Fay Wouk. One is a narrative elicited by asking the speaker to describe a film he has just seen about the adventures of a basket of pears (the "Pear Film", Chafe 1980) and the other is a reminiscence in which the speaker describes a childhood experience evoked by the pear film. I coded 242 clauses of Indonesian - 82 from the pear film, and 160 from the reminiscence. The speaker is Sumatran, and his first language is Toba Batak, but he has spoken Standard Indonesian since early childhood.

Malay and Indonesian, of course, are not really two different languages, but are distinguished today only for political reasons. For convenience of reference I will distinguish the two stages I have investigated by using the term "Malay" to refer to the older (Salleh) text, and the term "Indonesian" to refer to the modern (Manik) text. Malay/Indonesian has for centuries been involved in a sociolinguistically very complex situation, since it has been used as a lingua franca over a very large area for hundreds of years (there are Malay inscriptions in Sumatra dating from 700 AD). Thus there is really no such thing as a "representative" or "pure" data source at either of the stages which I am studying; variability along many parameters, including the factors influencing word order choice, is the rule rather than the exception. One might in fact say that Salleh is "typical" of Malay writers of his era in not having an established written standard (in the western sense of written - written for readers rather than for hearers) to conform to, and Manik is "typical" of Indonesian speakers of his era in speaking several other languages besides Indonesian, which possibly influence his speech. Identification of the factors which condition this variation must await a careful sociolinguistic study. There are differences between the two texts I looked at along many parameters, including geographical (one is from a Malay who lived in what is now Malaysia, the other from a Sumatran who has lived in Jakarta), social (one comes from the context of a public rendering of a life story, to an audience which isn't present, while the other is a recording of an informal narration delivered to friends and family), and channel (one is written and the other is spoken). Nonetheless both texts do represent outgrowths of a single linguistic tradition, and all of the differences between them are such as would lead us to expect that the Malay text represents a more conservative style, retaining features of earlier stages of the language, while the Indonesian text represents a later and more innovative stage; therefore their comparison may shed some light on the possible parameters of pragmatic change.

2. GRAMMAR

In this section I will outline those aspects of the morphology and syntax which are involved in word order alternation.

2.1 Terminology

I will use the terms A, P and S, more or less following Comrie (1981:64), as follows: the A is "that argument of the transitive construction which

correlates most highly with agent", P is "that one which correlates most highly with patient", and S is "the intransitive subject". However, since much work has shown that not all two-argument constructions are equally transitive, I have introduced a fourth term, C ("complement"), for the second argument of a construction which is not fully transitive. These arguments have syntactic and semantic properties distinct from patients: syntactically, they can't become the trigger of a patient-trigger construction; semantically, they are not highly affected. (This category is equivalent to what Rafferty (1982:14) calls "referent": "the non-volitional NP complement of a semi-transitive verb".) In determining the transitivity of clauses I have been guided by the morphology of the verb. Thus clauses in which the verb is prefixed with di- always have a (expressed or implied) P, clauses in which the verb is prefixed with ber- (and clauses which have a nonverb predicate, such as a noun phrase, prepositional phrase, or adjective phrase) always have an S and possibly a C, and clauses in which the verb is prefixed with meng- can have either A/(P) or S/(C) (Rafferty 1982, Dreyfuss 1978). Thus in a two-argument ber- clause such as[3]

(1) sengse' bernama Chia Asing (M:37)
teacher be:named Chia Asing
The teacher was named Chia Asing.

the first argument (sengse') is counted as an S and the second argument (Chia Asing) as a C.

I have also treated clausal arguments, which like complements of ber- verbs aren't affected, and which have similar syntactic restrictions, as Cs rather than as Ps. Thus in a clause like

(2) Saya nggak tahu apa itu pears dalam bahasa Indonesia. (I:1007)
I not know what that pears in language Indonesia
I don't know what "pears" is in Indonesian.

saya is treated as an S and apa itu pears dalam bahasa Indonesia is treated as a C.

2.2 The trigger system

I will use the neutral term "trigger" (introduced in Fox 1985) to refer to the syntactic role which linguists have variously called "topic", "focus", or "subject". There is a relationship between the semantic role of this NP and the morphology on the verb: if it is a P, the verb must have the prefix di-, ter- or a proclitic agent pronoun; if it's an A or an S the verb may not have these prefixes. Hence, this NP may be said to "trigger" the verb morphology. In Malay/Indonesian, clauses can be described as being either patient-trigger (PT) or actor-trigger (AT). A PT verb must have a patient, either expressed or clearly implied (the agent, if not expressed, may be vague or generic); an AT verb must have an A or an S, either expressed or clearly implied, but there may be no patient. Other syntactic properties of triggers will be discussed below.

2.3 Morphology

As mentioned above, PT verbs either have the prefix di-, the prefix ter- (which has the additional meaning of *accidental occurrence*), or a proclitic agent pronoun, which in Malay could be ku- *I* or kau- *you* but in Indonesian can be any pronoun (although it's still more common with first and second person). AT verbs take the prefix meng- (which usually is transitive but on some verbs is intransitive), the prefix ber- (which is always intransitive), or no prefix. There are some verbs (mostly intransitive) which never take a prefix, such as pergi *go*; in Modern Indonesian, meng- can always be omitted.

Both AT and PT transitive verbs can take the enclitic third person pronoun -nya; on an AT verb it represents the patient, on a PT verb the agent. The agents of PT clauses may optionally be marked with the preposition oleh, and the patients of AT clauses with akan.[4] The following examples illustrate AT and PT type clauses of various types in both Malay and Indonesian:

ACTOR TRIGGER:

(3) fakir pun hilir serta anak-anak fakir (M:76)
I PRT *go:downstream with children my*
I went downstream with my sons.

(4) jeruk jatuh, mangga jatuh, ... (I:75)
citrus fall mango fall
the oranges fall, the mangoes fall ...

(5) dan saya membawalah sepuluh orang panglima-panglima yang masyhur (M:2072)
and I MENG-*bring ten person chief* REL *famous*
and I took ten well-known chiefs

(6) saya pernah mencuri satu goni bawang (I:110)
I once MENG-*steal one sack onion*
I once stole a sack of onions

PATIENT TRIGGER:

(7) Tengku Nong diampun Kebawa Duli Yang Maha Mulia Tuanku kesalahnya (M:67)
T.N. DI-*forgive His Highness to:crime:his*
Tengku Nong was granted a pardon by His Highness

(8) biasanya ya orang dididik untuk bertanggung-jawab (I:100)
usually yes people DI-*teach to take:responsibility*
usually people are taught to take responsibility

(9) kadang-kadang orang terkejut (I:74)
sometimes people TER-*frighten*
sometimes people get startled (accidentally)

2.4 Zeros

These languages are quite free in their use of zero anaphora of various kinds, allowing agents, patients, and subjects to be freely omitted when they can be understood from the context. In addition to these pragmatic zeros there are obligatory zeros in relative clauses (marked by yang), complement clauses,

and adverbial clauses. If there is a missing argument which is definite and referential it is usually the trigger of its clause; the more tightly two clauses are linked the stricter this requirement is, so that it's obligatory with relatives, but merely preferred with clauses conjoined with a linker such as dan *and*. In the following examples, the clause containing a zero is capitalised.

(10) kamu tidak mau MENGIRIM SAYA KE SEKOLAH (I:133)
you not want MENG-*send me to school*
you don't want to send me to school

(11) tumbuh juga buah-buahan YANG MEMANG ENAK UNTUK DIMAKAN seperti di
grow also fruit REL *really delicious for* DI-*eat like at*
Samosir (I:1039)
Samosir
There is also fruit growing which is really good to eat like in Samosir.

(12) berkukuhlah pula Engku Ahmad itu kepada kubu musuh YANG
establish PRT *E.A. that at stockade enemy* REL
DIDAPATINYA (M:2047)
obtain:he
Engku Ahmad took up a position in an enemy stockade which he had captured.

(13) ya si petani ini memetik buah pearsnya DAN DIKUMPULKANNYA DALAM
yes PN *farmer this* MENG-*pick fruit pears:his and* DI-*gather:he in*
KERANJANG (I:1015)
basket
This farmer is picking pears and putting them into a basket.

(14) dan fakir jadikan satu peta jajahan Johor cukup dengan jajahan-jajahan
and I made one map territory Johor complete with territories
DAN DIHANTAR OLEH KEBAWAH DULI KE EROPAH LONDON (M:53)
and DI-*send by His Highness to Europe London*
and I made a map of Johor complete with all its territories, and (it) was sent by His Highness to London, Europe.

Since it is difficult to determine the position of a zero in a clause, and in this paper I am concerned with word order, in the statistics which follow I have counted only overt arguments. It is worth noticing, however, that a clause can be predicate-initial two ways: either by having a postpredicate trigger, or by having a zero trigger.

2.5 Word order

In the above examples the trigger, if present, was always prepredicate. This is the most frequent word order. However, the trigger can also occur postverbally:

(15) berbangkitlah suatu peperangan di dalam jajahan Muar (M:2001)
arise one war at inside territory Muar
A war broke out in the territory of Muar.

(16) malu saya kalau orang lain nanti dengarkan ini (I:18)
ashamed I if people other later hear this
I'll be ashamed if other people hear this later.

(17) dapat fakir pertolongan dari kerajaan di Teluk Belanga (M:30)
obtain I help from government at T.B.
I received help from the government at Teluk Belanga.

(18) dihantar fakir menjadi kerani di Iskander Puteri kerja (M:17)
DI-*send I* MENG-*become clerk at I.P.* *work*
I was sent to work as a clerk in Iskander Puteri.

This possibility, however, is limited to intransitives and to PT verbs: transitive meng- verbs must have a prepredicate trigger if the trigger is present.[5]

3. DIFFERENCES BETWEEN MALAY AND INDONESIAN

In this section I will discuss the statistical and absolute differences between Malay and Indonesian, with the aim of correlating changes in different parts of the linguistic system.

3.1 Word order change

As I mentioned in the introduction, Indonesian has substantially fewer clauses with Predicate-Trigger than Malay does. This fact can be expressed in several ways. In terms of the contrast between predicate-initial and argument-initial order (leaving adverbs, linkers, etc. out of consideration, and not counting relative clauses), a full 51% of Malay clauses are predicate-initial, as compared to 39% of the Indonesian clauses. However, this doesn't take into account the possibility of some arguments being zero. If only clauses in which the trigger is present are considered, and clauses with the presentative verb ada (which always precedes its trigger) aren't counted, 25% of Malay clauses are predicate-initial, as compared to 6% of Indonesian clauses.

These statistical differences are partly accounted for by an absolute difference between my Indonesian texts and my Malay ones. Indonesian is apparently adding a restriction to the use of the V T order: in the Indonesian texts I examined, only intransitive triggers occur postpredicately, while in Malay, this order is possible in all clause types (e.g. example (17) above), although more frequent in intransitive clauses. The Indonesian speaker accepts sentences with postpredicate triggers as grammatical, but apparently doesn't use them (or only uses them very rarely) in this speech style. This fact isn't enough to entirely account for the difference in the frequency of V T order, however. Even if we only consider intransitive clauses, 29% of Ss are postpredicate in Malay, as compared to only 8% in Indonesian.

It seems that to account for the frequency differences between Malay and Indonesian, then, we need to look for differences in the functions of V T word order. Looking at the characteristics of postpredicate triggers, we find that in both Malay and Indonesian, all of them have either been mentioned (either explicitly or by zero anaphora) in the immediately previous clause, are first or second person pronouns, or are first mentions. It is natural to assimilate the first two cases together; the speaker and the hearer, like just-mentioned

noun phrases, can be expected to be uppermost in the consciousness of the hearer. However, the third category hardly constitutes a natural class with the other two. Therefore I propose that there are two separate functions characteristic of the postpredicate trigger: the function of introducing new arguments, which one might call the "presentative" function, and the function of referring to established themes, which one might call the "continuity-marking" function, where "continuity" is used in the sense of Givón (1983).

3.1.1 "Presentative" function

As we have already seen, V T order is obligatory for the verb ada, whose use is specialised for introducing new referents.[6] However, this order also occurs with other intransitive predicates (often verbs of motion) which introduce new participants (postpredicate triggers are upper case in these examples):

(19) datang PENGHULU RAHMAT pada fakir di hulu sungai Batu Pahat (M:072)
come chief Rahmat to me at head river Batu Pahat
Chief Rahmat came to me at the upper reaches of the river Batu Pahat.

(20) pada waktu yang sama tumbuh juga BUAH-BUAHAN (I:1039)
at time REL *same grow also fruit*
at the same time, there is also fruit growing

It is unsurprising that we should find this order serving this function; in many languages, even "rigid" SV languages like English, we get VS order in precisely these kinds of contexts.

3.1.2 "Continuous themes"

It has been suggested by Givón (1983) that word order should correlate cross-linguistically with "thematic continuity", that is, the position of an argument should reflect how recently it has been mentioned; in particular, post-predicate position is held to be more continuous than prepredicate position for languages with "free" word order. Verhaar (forthcoming) suggests that this is indeed the main conditioning factor for Indonesian word order. My data support this suggestion, as the following examples show:

(21) telah dapatlah setengah kubu-kubu musuh ke tangan Engku Ahmad itu
ASP *obtain half stockade enemy to hand E.A. that*

berkukuhlah pula ENGKU AHMAD ITU kepada kubu musuh yang didapatinya
establish PRT *E.A. that at stockade enemy* REL DI-*obtain:he*

pada suatu tempat yang bernama Lubuk Bandan. (M:2046)
at one place REL *named Lubuk Bandan*
A number of enemy stockades had fallen into the hands of this Engku Ahmad; that Engku Ahmad took up a position in an enemy stockade he had captured in a place called Lubuk Bandan.

(22) Terus dia angkat satu keranjang ke dia punya sepeda, lari pergi DIA,
then he puts one basket to he POSS *bike run go he*

bawa mencuri. (I:1074)
carry steal
then he puts one basket on his bike, he runs away, carries it off, steals it.

So far, however, we haven't accomplished the original goal of differentiating Malay and Indonesian according to the function of prepredicate position. A difference shows up, however, if we look at how well the presentative and continuity functions predict postpredicate position. While both languages have many exceptions to the principle that first mentions and continuous themes will be postpredicate, Malay has considerably fewer than Indonesian does, as the following table shows. In table 1, the number of postpredicate triggers is given as a percentage of the total number of first mentions, and as a percentage of the total number of continuous themes for each language. In table 2, the number of postpredicate triggers of intransitive clauses (Ss) is given as a percentage of Ss which are first mentions, and as a percentage of Ss which are continuous themes.

Table 1: postpredicate triggers (all roles)

	Malay	Indonesian
first mentions:	72%	42%
continuous themes:	24%	19%

Table 2: postpredicate Ss

	Malay	Indonesian
first mentions:	42%	16%
continuous themes:	27%	7%

The fact that there are so many exceptions in each case is undoubtably due to the fact that there's a lot more to "presentative function" than being a first mention, and there's a lot more to "continuous theme" than being a first person pronoun in a personal narrative or having been mentioned in the previous clause; getting the working definitions of these concepts fine-tuned enough to work all the time will involve arriving at a better understanding of the episodic structure of my texts, which in turn will involve having dealt with longer portions of them. In the meantime, I will continue to use the terms "continuous theme" and "presentative" on the assumption that it really is these functions, or something close, that "first person or mentioned in the immediately preceding clause" and "first mention" are getting at.

But what are we to make of the fact that in Indonesian the figures are so much lower? What this suggests to me is that there is another principle which favours prepredicate position for triggers competing with the one which favours postpredicate position for certain triggers. In the following section I will outline what that principle might be.

3.2 Morphological change

In addition to the change in the frequency of V T word order mentioned above, there have also been some morphological changes which have as one result an increase in the frequency of sentences in which As precede verbs.

3.2.1 Spread of procliticisation

As remarked above, there has been a change in what pronouns could take the place of the di- prefix in PT constructions. In older Malay the possibilities were limited to ku- *I* and kau- *you*(sg.); both of these are informal forms, titles and various other forms of address being preferred, and don't occur at all in my texts. The following example is taken from Winstedt's grammar (Winstedt 1914).

(23) hai Ruana KAU-katakanlah yang kehendak hatimu itu, KU-dengar
oh Ravana 2sg-*tell:*IMP REL *desire heart:your that,* 1sg-*hear*
Tell, O Ravana, the desire of your heart, that I may hear it.

In Modern Indonesian, any pronominal form (including derived second person forms such as tuan (= *lord*), saudara (= *sibling*)) can be used in this fashion. They are written as separate words, but are considered to be procliticised because they cannot receive stress or be separated from the verb by an element such as an auxiliary: in the following example (taken from Wolff 1980), harus *must* could not come between the agent saudara and the verb.

(24) apa lagi yang masih harus saudara cari?
what more REL *still must you seek*
What else do you still have to look for?

The result of this change is that there are many more clauses with A V word order in Indonesian than in Malay, as A+V clauses replace di-V A clauses.

3.2.2 Loss of trigger morphology

In Modern Spoken Indonesian there are many transitive verbs which would have had a prefix in older Malay (and still do in the standard language) which have no prefix. Grammarians have generally assumed that these are meng- forms with the prefix "dropped" unless there is evidence of procliticisation (cf. Wolff 1980), although the syntactic properties of these verbs aren't altogether clear.[7] These forms are in variation with forms in which the meng- is present; this variation is apparently stylistically rather than syntactically conditioned, with the meng- form being more formal (Wolff 1980). Variation between prefixed and non-prefixed forms is illustrated in the following excerpt:

(25) you curi jeruknya bawa kesini, memasukkan kesini begini, masukkan
you steal citrus:his carry here MENG-*put:in here thus put:in*
kesini, mencuri, ya? (I:57)
here MENG-*steal, yes*
You steal his citrus, carry (it) to here, put (it) in here like this, put (it) in here, steal (it), see?

This example comes from a demonstration of how to steal fruit by hiding it in your sarong. The verbs curi *steal* and masukkan *put in* both occur with and without prefixes.

This change, taken together with the spread of agent procliticisation, leads to the occurrence of clauses which could either be described as PT with procliticised A, or as AT with deleted meng-:

(26) terus dia lihat pears (I:1069)
then he see pears
then he sees the pears

This kind of example isn't at all rare; there are 11 cases in my Indonesian data which have this kind of indeterminacy (about 5% of all clauses). There is no ambiguity that arises from these cases, however; in any case the preverbal NP is the A, and the postverbal one is the P. What seems to be happening is that the trigger system is being lost in favour of a simple AVP word order strategy.

Another difference between the Malay and Indonesian texts which contributes to this reanalysis is that in the Indonesian texts, there are no cases of independent full NP As in di- clauses. While the speaker accepts this clause type as grammatical, it is rare enough that it doesn't occur in my data. There are enclitic agents (as in example (13) above), and cases of bare verb stems in which the syntactic constraints on relativisation force us to the conclusion that we have a proclitic agent:

(27) apa saja yang kita minta mereka harus kasih (I:120)
whatever REL *we ask they must give*
whatever we ask, they must give

(In this sentence, the relative clause must be PT, since the P is extracted, so kita must be proclitic.) However, you don't find nonclitic As such as occur in Malay clauses like example (7) above, or like the following:

(28) diambil Enci' Abu Bakar bakal raja (M:7)
DI-*take Encik Abu Bakar future ruler*
(I) was taken by the Encik Abu Bakar, the heir apparent.

Thus this construction is apparently being used much like an English passive, when required by the syntax due to a zero argument, as in examples (12-14) above, or when the agent is either obvious from context, or unknown, or not important, or some combination of these, as in (18) (cf. Thompson, forthcoming, on the function of the English passive). The result is that 100% of the non-enclitic As in my texts are preverbal. In fact, it seems that occasionally you even get relative clauses which ought to be PT, because of the relativisation of the patient, but which aren't, with the result (again) that you get preverbal agents even in these clauses. This doesn't occur in my data, but other authors cite the following two examples:

(29) Ia membawa buku yang baru ia membeli dari toko (Becker 1980)
he MENG-*bring book* REL *just he* MENG-*buy from store*
He brought the book he just bought from the store.

(30) Semua hal yang dia tidak lihat harus saya urus (Verhaar, forthcoming)
all thing REL *she not see must I arrange*
All the things that she does not see I have to arrange.

(The relative in (30) must be AT, since tidak intervenes between the A and the verb.)

4. VERB-INITIAL TO ARGUMENT-INITIAL

All of these developments have a common outcome, namely that PT clauses are getting rarer (from 14% of all clauses in Malay to only 8% in Indonesian), and preverbal As are becoming the norm. Thus the word order shift between Early Modern Malay and Modern Spoken Indonesian can be seen as primarily an increase in T V order rather than as a decrease in V T order - a continuation of the trend which has been going on since the hypothesised verb-initial pre-Malay stage. Of course, the factors conditioning the placement of the A are only relevant to transitive clauses; therefore the S has retained its freedom of position to a greater extent.

These developments also have another thing in common: they make Indonesian look more like English. The loss of trigger morphology on the verb, the increase of AVP word order, the reduction of PT clauses with expressed As, the restriction of VT order to intransitives - all make the Indonesian "trigger" system look superficially much more like an English-type subject system, in at least three respects. First, A and S tend to pattern together with respect to word order, as opposed to P. In the earlier system this wasn't the case, due to greater variability of order for all three rules. Second, the cases in which you do get a preverbal P involve morphological marking on the verb (the di- prefix, which shows up when there is no agent to "procliticise"), whereas verbs which don't have preverbal P are increasingly unprefixed, hence unmarked in the morphological sense. Third, the increasing restriction of prefixed PT clauses to agentless contexts greatly reduces their frequency, and thus it seems that it is becoming more "marked" in the non-morphological sense as well. One of the major differences between the Austronesian type trigger system and a voice system is that in the trigger system there is a symmetry between the various trigger choices in the verb morphology and the case marking morphology. Thus, in an English transitive active clause there are two direct arguments and no marking on the verb, while in a passive clause there is one direct argument and marking on the verb; in many Austronesian languages, however, A and P are oblique (as in Tagalog) or direct (as in Malay) alike in AT and PT clauses, and both clause types have verbal marking. Indonesian, however, seems to be losing this symmetry.

This shift may already have begun in the Early Modern Malay text I looked at. It is certainly the case that "high" (literary) Malay texts from a generation earlier that have been studied by researchers (such as the *Hikayat Abdullah*, the primary source for Hopper 1983, and one of the sources analysed in Rafferty 1984) show a higher percentage of V T order, with this order in some texts found in the majority of clauses. There is evidence in these earlier texts that V T order correlates best with "eventive" narrative (Hopper 1983). While the distribution of word order in the texts I looked at isn't satisfactorily characterised by this hypothesis, it may be that the situation described by Hopper represents a developmentally earlier state of affairs. With this in mind, we can hypothesise the following sequence of events:

1. "Pre-Malay" was a predominantly predicate-initial language which had some preverbal arguments. These arguments were related to important themes of the discourse, as is the case in Toba Batak and Tagalog, sister languages of Malay. Being associated with discontinuous themes in particular, as is the case in Tagalog, they tended to occur at the beginnings of episodes (when a set of participants needed to be introduced or reintroduced).

2. Because of this correlation between episode-initialness and argument-verb order, this order became associated more generally with a scene-setting function, while verb-argument order became associated with an episode-medial "eventive" function. Thus argument-verb order became more widespread; in particular, besides the function of recalling expired themes, it added the function of providing scene-setting or "backgrounded" information (extrapolated from Hopper 1983; his "active" and "passive" clause types are both usually argument-initial, and are said to share the "background" function). This is the situation which obtains in the *Hikayat Abdullah*, Hopper's source.

3. Meanwhile, changes were taking place in spoken Malay. Among these was a restriction of V T order to the presentative function and the function of coding continuous themes. The latter has a fairly natural connection with eventive discourse, since it is within an episode rather than at its beginning that you are likely to get continuity of participants; thus this can be seen as a matter of reanalysis also. Salleh's text reflects this state of affairs, since it apparently uses a style further from the classical literary style and closer to the spoken than Abdullah, and it is also chronologically later.

4. Subsequent to this stage, the morphological changes discussed in this paper occurred, with the results for word order which have been discussed.

Whether or not the details of this hypothesis are true, one result emerges: it would seem that the drift from verb-initial to argument-initial in Malay is not a unitary phenomenon with a single motivating force, but rather involved different kinds of reanalysis at different times. In particular, the morphological changes which have clearly resulted in a much higher proportion of argument-verb clauses in Indonesian had not occurred in Salleh's Malay, and yet there was clearly a stage before Salleh's time which was more predicate-initial than Salleh's language is. The change from that stage to Salleh's stage must have been rather different in nature from the change between Salleh and Modern Indonesian. If there was not a common motivating factor, why should changes at different points in the history of the language lead to the same result, namely, more argument-initial syntax? The most obvious explanation is that contact over the centuries with argument-initial languages has had this effect; a useful test of this hypothesis would be the examination of a parallel shift in a language which was not in a comparable situation of contact, for example, Old Javanese. This is a fruitful area for further research.

NOTES

1. In Malay/Indonesian, as in many Austronesian languages, many sentence types do not have a verb, but rather a predicate which is a noun phrase, prepositional phrase, or adjective. Thus I will use the term "predicate" rather than "verb" when discussing word order, and limit the term "verb" to words which are morphologically verbs. I will use "V", however, to abbreviate both "predicate" and "verb".

2. Grateful acknowledgement is hereby made to Ellen Rafferty and to Fay Wouk, without whose patient help, instruction, and advice this paper would not have been written. Any deficiencies are purely my own.

3. The notation following each example gives the language that the example comes from (M = Malay, I = Indonesian), and then the number of the clause in my records.

4. Since agents and patients marked thus have different syntactic properties than arguments not so marked (oleh phrases have more freedom of position, akan phrases have less) and probably also different discourse properties (Hopper 1983 suggests that oleh decreases transitivity, and akan increases it), I have coded them as obliques rather than as A and P.

5. This restriction is referred to in the grammars, and holds for my data; Rafferty (personal communication) has worked on some earlier texts which contain occasional meng- verbs with postverbal trigger.

6. Ada occurs with other orders as an emphatic auxiliary, and to predicate possession and location, but these can be considered separate usages.

7. Wouk (1983) provides some evidence from her elicitation that they behave like AT verbs with respect to relativisation of the A.

REFERENCES

BECKER, A.L. and Umar WIRASNO

1980 On the nature of syntactic change in Bahasa Indonesia. In Paz Naylor, ed. *Austronesian Studies: Papers from the Second Eastern Conference on Austronesian Languages*, 95-102. Michigan Papers on South and Southeast Asia 15.

CHAFE, Wallace, ed.

1980 *The Pear Stories: cognitive, cultural and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

COMRIE, Bernard

1981 *Language universals and linguistic typology*. Chicago: University of Chicago Press.

CREIDER, Chet A.

n.d. Thematization and word order. MS.

DREYFUSS, J.V.

1978 *Men-, di-* and *ber-*: three analyses. In John W.M. Verhaar, ed. *Miscellaneous studies in Indonesian and languages in Indonesia*, part V, 1-6. NUSA 6. Jakarta: Universitas Atma Jaya.

FOX, Barbara A.

1984 Discourse structure and morpho-syntactic markedness. Talk given at USC, February 1984.

1985 Word order inversion and discourse continuity in Tagalog. In Talmy Givón, ed. *Text* 5/1-2:39-54.

GIVÓN, Talmy

1983 Topic continuity in discourse: an introduction. In Talmy Givón, ed. *Topic continuity in discourse: a quantitative cross-language study*, 1-43. Amsterdam: John Benjamins.

HOPPER, Paul J.

1979 Some discourse origins of ergativity. *Working Papers in Linguistics University of Hawaii* 11:137-153.

1983 Ergative, passive and active in Malay narrative. In Flora Klein-Andreu, ed. *Discourse perspectives on syntax*, 67-88. New York: Academic Press.

MacDONALD, R. Ross and Soenjono DARJOWIDJODJO

1967 *A student's reference grammar of modern formal Indonesian*. Washington: Georgetown University Press.

POEDJOSOEDARMO, Soepomo

1982 *Javanese influence on Indonesian*. (Ph.D. thesis, Cornell, 1970). *PL*, D-38.

RAFFERTY, Ellen

1982 *Discourse structures of the Chinese Indonesian of Malang*. NUSA 12. Jakarta: Universitas Atma Jaya.

1984a Languages of the Chinese of Java - an historical review. *Journal of Asian Studies* 43/2:247-272.

1984b Word order in High and Low Malay of the late 19th century. Paper presented at the Conference on Linguistic Coding of Discourse Relations, University of Oregon, June 1-3 1984.

STAROSTA, Stanley, Andrew K. PAWLEY and Lawrence A. REID

1981 The evolution of focus in Austronesian. In S.A. Wurm and Lois Carrington, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 145-170. *PL*, C-75.

SWEENEY, Amin

1980a *Reputations live on: an early Malay autobiography*. Berkeley: University of California press.

SWEENEY, Amin, ed.

1980b *The Tarikh Datu' Bentara Luar Johor*. University of California at Berkeley: Center for South and Southeast Asian Studies.

THOMPSON, Sandra A.

forthcoming The passive in English: a discourse perspective. In Robert Channon and Linda Shockey, eds *In honour of Ilse Lehiste: Ilse Lehiste suhendusteos.* Dordrecht: Foris.

VERHAAR, John W.M.

forthcoming Syntactic ergativity in contemporary Indonesian. To appear in Richard McGinn, ed. *Proceedings of the Third Eastern Conference on Austronesian Linguistics*.

WINSTEDT, R.O.

1914 *Malay grammar*. Oxford: Clarendon Press.

1939 *A history of classical Malay literature*. Monographs on Malay Subjects 5. *Journal of the Royal Asiatic Society, Malayan Branch*, 31/3.

WOLFF, John U.

1980 *Beginning Indonesian, parts one and two*. Ithaca: Cornell University, Southeast Asia Program.

WOUK, Fay

1983 The syntax of bare verb stems in Indonesian. MS. UCLA.

THE ASPECTUAL PARTICLE *NUNGA* IN TOBA BATAK

Susan Mordechay

Among the diversity of predicate oriented particles in Toba Batak (TB), nunga, as well as a few others, would be classified as an aspectual particle. Such particles have by now achieved the status of legitimate objects of linguistic research and an increasing number of studies have been fully dedicated to their description and analysis.[1] This paper, in like manner, is fully devoted to the description and analysis of the variety of uses of nunga in TB.

Existing grammars of TB have paid a certain amount of attention to nunga along with the other predicate particles. Nababan (1966, 1981), for example, considers it to be an "explicit completive" marker, and Percival (1964, 1981) translates it as *already*. But while both these characterisations are well founded, as we will see, more needs to be said if a thorough understanding of nunga is sought. First, there immediately arises the question of whether the two characterisations are at all compatible with each other, and if they are, how? The second, and more significant question is, to what extent is each one an adequate characterisation of nunga? From the outset we cannot be sure that a category such as "completive" would mean exactly the same thing in the description of TB, as it does, for example, in the description of English. It is possible that the range of phenomena covered by "completive" in TB would overlap only partly with that covered by that category in English. Analogously, it is possible that while nunga is correctly translated as *already* in some of its uses, such translation is inadequate in other of its uses.

The first part of the paper examines the possibility of viewing nunga as a marker of the Perfect. The Perfect category was chosen as an initial approximation for a definition of the function of nunga as nunga clauses were obtained whenever English Perfect clauses were elicited in isolation from a speaker of TB.[2] Since I use a definition of the Perfect category which properly subsumes the Completive category, the discussion in this part implicitly touches the question of the latter category as well. The second part of the paper examines occurrences of nunga, most of which are taken from actual texts, which are not naturally interpreted as marking the Perfect. In some cases the category Perfect is shown to be too narrow to capture further aspects of the use of the particle, in other cases, it is too broad. I consider here the possibility of translating nunga as *already*, and show that while the two particles do indeed have a lot in common, the distribution of *already* is too restricted to render a full account of nunga uses. However, the properties these two particles share form the basis for a general definition which, I think, accounts for the majority of the functions of nunga. Nunga is shown to indicate that the situation expressed by the rest of the clause is a *stage* in a *schematic process*, where a schematic process

---

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics,* 113-132. *Pacific Linguistics*, C-93, 1986.

is a series of ordered situations conceptualised into a single idea. A further generalisation of the notion of a schematic process provides a characterisation of nunga which covers *non-temporal* uses of the particle as well.

1. NUNGA AS A PERFECT MARKER

The grammatical category of the Perfect[3] covers various types of uses, which can be illustrated by the following English sentences:[4]

(1) a. John has written 3/4 of the book.
b. When we arrived, everybody had (already) been evacuated.
c. Have you (ever) been to China?
d. The soup is hot now.
e. Rita has lived here for ten years.
f. I'll have finished by the time you get back.

In each of these sentences a state is presented as prevailing at some reference time and as relating to a situation[5] at a previous time. By *reference time*[6] we mean the time of the speech event or the "now" as in (1a,c,d and e) or else a time given in the context as in (1b) and (1f). At the reference time, there is generally a *state* rather than any other kind of situation, which exists, or which has been realised, whereas the earlier situation can be of any type - an event, an activity or yet another state.

By definition, it is the state at the reference time rather than the previous situation which is the focus of attention of the Perfect sentence. Although for many types of the Perfect (for instance (1a,b,c and f)), the earlier event is the one mentioned explicitly in the sentence, it is not mentioned "for its own sake", but rather for the sake of its consequences or its delayed effects. In other types, for example (1d) and (1e), the current state is the one explicitly mentioned in the sentence, and the sentence, due to its Perfect form, implies a previous situation. In any case, whatever the form of its appearance in a Perfect sentence - verbally present or not - the current state is "what the sentence is all about". By virtue of the use of the Perfect form, however, the current state is *presented* as essentially and inseparably connected with a previous situation.

Given a certain situation, then, there seems to be a choice as to how to present it. A speaker may choose to present an earlier event, for example, as temporally separated and not included in the period s/he identifies as the present, or s/he may choose to present the same event as part of his or her present. In English, for example, the speaker would use the preterite in the first case, and in the second, the Present Perfect.

But the speaker's choice in presenting situations is not independent of the overall contextual conditions. If, for instance, the embedding context concerns a certain present situation, a speaker would typically present a prior event in the Present Perfect. A conversation dealing with the immediate situation the discussants are in is, thus, a likely context for the Present Perfect in English. And in narratives, where certain events are presented as linearly ordered, forming the primary time axis of the narrative, an event would typically be presented in the Perfect if it is not this event but a closely related subsequent situation which lies on the primary time axis of the narrative.

In the light of these remarks it would seem justified to view nunga as marking the Perfect if it can be demonstrated that its presence in a sentence

has the effect of relating the current state to a previous situation whereas, in its absence, a similar sentence is neutral with respect to such a relation. In what immediately follows we will see that indeed that is the case.

(2)a. Sahat si John.
arrive PN[7]
John arrived.

b. NUNGA sahat si John.
PF *arrive*
John has arrived.

In a conversational context, both (2a) and (2b) may be interpreted in two ways according as the time of the mentioned event - John's arrival - is supposedly in the past or in the present. For the moment we will be concerned only with the former case, the latter will be discussed later on. (2a) and (2b), then, may refer to the same past event, but they do so in different ways: (2a) simply says that John arrived; nothing more, nothing less. (2b) on the other hand, refers also to the present situation, and actually can be glossed as *now, John has arrived.* (2a) would typically be used as part of a narration of past events and normally will be accompanied by the particle ma,[8] as in ... sahat ma si John, meaning ... *then John arrived.* (2b) on the other hand would typically be used when the present situation, the "now", is discussed, and when John's arrival has some particular relevance to that situation.

Similarly, both (3a) and (3b) may refer to a past event, but while (3a) is a simple report of reading "the book", the reference made to that activity in (3b) is meant to saying something about "now". For example, in a book store, an offer might be made to the customer to buy a certain book which the latter does not care to buy. It would be appropriate then to use (3b) to reject that offer. The discourse really concerns the situation the discussants are "currently" in, and the use of nunga here indicates that this situation is viewed as an effect of the earlier one of "reading the book".

(3)a. Hujaha buku on.
I-read book that
I read that book.

b. NUNGA hujaha buku on.
I've (already) read that book.

(4)a. Udan.
rain
It's raining.

b. NUNGA udan be.[9]
Now, it's raining.

In (4), both sentences make a statement about the weather at the moment of speech. (4a) presents a plain picture of the current situation and by itself lacks any concern with any other situation. It might be uttered during a phone conversation as an answer to "how is the weather over there?" or "how come you are at home at this time (of the day)?" (4b), on the other hand, indicates that the current rain is somehow new, and therefore is a change from a previous situation. It might be uttered at the beginning of the rainy season as a response to the first question above, or if it has been cloudy for a while, or even if in reality it has been raining for a while but this is the first moment the speaker has noticed the rain.

The sentencesin (5) may depict an on-going situation contemporaneous with the time of speech.

(5)a. Manjaha ibana
read he
He is reading.

b. NUNGA manjaha ibana.
Now, he is reading.

Again, (5a) is straightforward and neutral with respect to the onset of the activity, but (5b) is not neutral, and implies a change. It might be said about a child just having mastered reading, or when someone has been urged for a while to do some reading and has finally yielded to the pressure.

In all these cases it is nunga which induces the relation felt between one state of affairs and some other earlier one. But a notable distinction in terms of aspectual properties exists between (2) and (3) on the one hand, and (4) and (5) on the other hand. Namely, (2) and (3) contain a *perfective* predicate each, and (4) and (5), an *imperfective* one. That is, the former sentences present a situation as a single whole, viewed in its totality, with no hint as to any internal temporal complexity. The latter ones, on the other hand, present a situation as something on-going at the reference time, and allude to some internal temporal structure (cf. Comrie 1976). This distinction in the aspect of the predicate results in a difference in the kind of implication produced by the presence of nunga. When the predicate is perfective,[10] the effect of nunga is to produce an indirect reference to an unexpressed "current" state which is later than the depicted event. But when the predicate is imperfective, expressing a situation existing "currently", the effect of nunga is an implied indication of a different state of affairs which is earlier than the explicitly expressed one. Further examples of the interaction of nunga with these two classes of predicates are presented below separately.

1.1 Nunga with perfective predicates

(6) NUNGA sae di ahu.
PF *finish* PRT *I*
I've finished.

(6) can be uttered in a dinner situation where one is offered some more food by the host and is trying to politely refuse. (6) here can mean "Thanks, I've had enough", "Really, I cannot eat a thing more ..." or "I don't want any more ...". The speaker-eater is in fact commenting on his or her present situation relating it to the earlier one in which s/he hadn't yet finished eating.

(7) NUNGA hutonggor (be) pillim on.
I-see PRT *film this*
I've already seen this movie.

(7) might be uttered as a response to a suggestion to go to a certain movie. It might stand for either a negative or a positive response to the suggestion: "I've seen the movie, so I don't want to go", or " ..., O.K. I'll see it again". In both cases, the situation under discussion is again the present situation and nunga is the means of bringing into the discussion a relevant fact concerning a

past event. Stripped of nunga (and be), (7) would still express the same proposition but would by itself be neutral with respect to the effect of the expressed fact on the present situation. It - (7) minus nunga - might be used in the above context, but a more plausible context for it would be an account of a sequence of past events of which it would be just one in the sequence. In the latter context, the sentence including nunga would be inadequate.

To further demonstrate that nunga would be used to connect a past situation with the present, the following two examples are augmented with short contexts in Toba Batak. In (8), the nunga clause is immediately succeeded by a remark about the current condition of the speaker, and in (9) it is immediately preceded by a request for help. Additional explanation seems redundant here as the glosses speak for themselves:

(8) NUNGA pagodanghu ahu modom: alani ndang tarpapodom ahu be.
too-much I sleep therefore no can-sleep I PRT
I've slept too much, therefore I can't sleep any more.

(9) Ise do boi mangarupi ahu mangantusi on? Si Bob, alana NUNGA diparsiarjari
who PRT *can help I understand this* PN *because* PF *study*
imana sudena buku i.
he all book the
Who can help me understand this? Bob, because he has already studied the whole book.

Next are two examples from recorded texts:

(10) NUNGA hudok nangkin, adong ma i rupani ruma bona.
I-said earlier there-is PRT *this is-called*
As I've said earlier, this is what is called ruma bona.
(*Jabu Halak Batak* 46)

(11) Ia orde baru, saonari itiop angka halak jenderal do. Jala on NUNGA
as-for order new present held PLR *people general* PRT *and this*
dipabotohon nasida on tu angka uarga nagara.
let-know they this to PLR *member-nation*
As for the present new order, it is controlled by generals. And they have already let the citizens know it.
(*Conditions in Indonesia* 19-20)

The initial part of (10) is a typical clause and is similar in its use to its English gloss *As I've said earlier* .... It is part of a description of a traditional Batak house fashioned as a tour through the different storeys and rooms of such a house. It contains both a statement which is a repetition of an earlier statement and a meta-linguistic remark - the nunga clause - commenting on that statement and acknowledging its repetitiousness. The immediate context surrounding (10) is an interlude in the general flow of the description, the narrator having encountered difficulties in explicating the significance of a certain structure. This interlude contains other meta-linguistic remarks such as "It's complicated" and "this is impossible to be explained" and following (10), two more echos of earlier comments. It seems appropriate, therefore, to view it as concerning primarily the current status of the ongoing description. Being part of that interlude, the nunga clause in (10) is likewise commenting actually about the present, while making an explicit reference to the past.

(11) is taken from a text which discusses the economic and political conditions in Indonesia. As its first sentence explicitly indicates, it is about the present "new" order, controlled by the current generals' regime. The nunga clause assures the reader that the fact that the country is controlled presently by generals is *by now* clear to the public, presumably through certain acts performed by the generals.

All the examples so far have been in the present tense and concern the time of speech. But, as with Perfects in other languages, the reference time of nunga need not be the present; another tense may be established in the context and the reference time may belong to that tense. Thus in (12a), the reference time is in the past, in (12b), it is in the future, and in (12c) it is a general hypothetical time:

(12)a. Tingki sahat ahu di pesta nabodari, dipaboa si John tu ahu naung
when arrive I to party last-night tell PN *to I* COMP+PF

laho si Mary.
go PN
When I came to the party last night, John told me that Mary had (already) left.

b. Molo dung sahat ahu di pesta marsogot, NUNGA be borhat si Mary.
if after arrive I to party tomorrow come PN
When I get to the party tomorrow, Mary will have already arrived.

c. Jala, molo NUNGA rupani denggan-denggan dihatai, jala natua-tua NUNGA
and if as-we-say good discuss and parents

nagihut di hata ni anak, ba adong ma i na ginoarnna manungkun
follow word of son PRT *there-is* PRT *that* COMP *called price*

boru.
woman
And if, as we say, it has been well discussed and the parents already follow the words of the son, then there is the so-called bride-price.
(*Pangolion* 7)

(12c) is taken from a text describing the sense of the traditional Batak marriage process. The text is narrated in definitional, generic terms, hence references to time mostly represent general rather than particular or even habitual times. This example will be discussed at greater length later, and is mentioned here solely for the sake of showing it is possible for nunga to occur in such abstract contexts.

Thus we have seen that with a perfective predicate, nunga is used to connect the situation depicted by the predicate with another situation which is later in time and which is the focus of attention of the discourse. Before we proceed it should be noted that, as in the case of English and Perfect markers in other languages, nunga by itself does not provide any idea as to the length of the time span which includes both the time of the expressed prior situation and the current state. In (7), for instance, which deals with having seen a certain movie, that time span includes the speech time and the whole previous life time of the speaker. But this clearly could not be the case in (8) which deals with the ability (or inability) to sleep at a certain time. What matters when the length of time is of interest is the lexical semantics of the predicate, "real world knowledge", and the context of utterance. But I will not pursue this issue further.

1.2 Nunga with imperfective predicates

As mentioned earlier, the imperfective is the temporal aspect associated with predicates presenting states, on-going activities and habitual or iterative situations. In Batak such predicates are adjectival or nominal, as in other languages, or constructed with e.g. the mang- and ni- forms of the verb.[11] When nunga combines with such predicates it adds the implication that the situation expressed by the predicate as present at the reference time has not existed at some earlier time and, thus, some change must have occurred.

Consider the following minimal pairs:

(13)a. Mohop sup i.
hot soup the
The soup is hot.

b. NUNGA mohop sup i.
The soup is hot now.

(14)a. Tung mansai paargahu sipanganon.
so so expensive food
Food is so expensive!

b. NUNGA tung mansai paargahu sipanganon.
Food has become so expensive!

Each of the a-sentences in (13) and (14) asserts merely that its subject has a certain property at the reference time. These sentences say nothing as far as any other state of their subjects at any other time is concerned. The b-sentences, on the other hand, do exactly that. In both, a state at a different time is implied which is different with respect to the mentioned property from the one depicted in the sentence. Thus (13b) implies that at an earlier point the soup was not hot, and (14b) implies that food was not always as expensive as it is "now".

Similarly, (15a) implies a state where John was not (yet) a teacher, and (15b), a state where the baby was not (yet) able to eat fish:

(15)a. NUNGA guru si John.
John's now a teacher.

b. NUNGA mangallang dengke si unsok.
eat fish PN *child*
The child/baby already eats fish.

Of course, certain discourse conditions must exist in order for such implications to be natural. Thus an appropriate situation for (13b) might be one where the soup is warming up, expected to become hot and finally indeed does become hot. And a type of discourse situation which could elicit statements as those in (15) is one where, as Li, Thompson and Thompson (1982) write, considering the Mandarin Perfect particle Le, requests are made by the audience to be "brought up to date about a person, a certain event, or a certain situation". In such contexts, the speaker could respond with a nunga sentence perceiving the state s/he describes to be new to his/her audience, a change from a previous state s/he assumes the audience is aware of. That is, in using a nunga sentence in such contexts the speaker ties the current new state with the one presumed to be known to the listener.

In a similar sense, a speaker may use nunga when a situation is new to him although it is not totally new objectively. Thus (16) could be uttered when in reality the time is a few minutes past eight o'clock but this is the first moment the speaker realises it is indeed so late.

(16) NUNGA jom aula!
hour eight
Wow, it's eight o'clock (already)!

Note that in (13) through (16) the reference time has not been explicitly given and thus has been assumed to be the present time or "now". As was the case with the perfective predicates, this is not necessarily so. The next two examples illustrate reference times established in the discourse:

(17)a. Ai jom lima manongot ba NUNGA antong pinagagt horbo i.
as-for hour five morning PRT PF PRT *be-fed buffalo the*
At five o'clock in the morning, the buffalo is fed (already).

b. Ai jom lima manogot ba NUNGA antong dunggo iba.
as-for hour five morning PRT *wake-up self*
At five o'clock in the morning I'm awake already.
(*Daily Life in Samosir* 3)

(17a,b) are almost identical and both are present here in order to illustrate the different types of predicate that may combine with nunga. Both refer to an iterative rather than a specific situation, but this fact cannot be inferred unless these sentences are viewed in their broader context. (17b) is part of a text describing the daily life in a Batak village and hence depicts an habitual, daily situation. (17a), though an elicited sentence, was modelled after (17b) and is habitual as well. Both assert the daily existence of a certain state of affairs - "I am awake" or "the buffalo is fed" - at five o'clock in the morning, and relate it to an earlier situation in which "I am not awake" or "the buffalo is not fed" which is the situation presumably assumed by the audience.

(18) NUNGA jom sada tingki dunggo imana.
hour one when wake-up she
It's (already) one o'clock when she wakes up.

(18) is similar to (17) in that it too may be interpreted as depicting an habitual situation. But it is different from it in that here the clock time expression denotes that which is predicated on the reference time whereas in (17) the clock time denotes the reference time itself. In (18), the latter is expressed by the clause "when she wakes up".

Here are two more examples from texts:

(19) Itor ba irupi ma imana paunehon pears i tu karanjang i muse
then PRT *help* PRT *he straighten pears the to basket the again*

naung rage di tano i ...
COMP+PF *spread in ground the*
Then they helped him set back into the basket the pears that were (already) scattered on the ground.

(20) Jala somalna di jaman moderen saonari, upacara mangadati on NUNGA
and usually in era modern present ceremony this

ditinggalhon, alana ndang sanggap halak alana paargahu.
be-left-behind because not capable people because too-expensive
And usually in modern times, this mangadati ceremony has been left behind because people cannot afford it because it is too expensive.
(*Pangolion* 50)

(19) is taken from a narrative called the Pear Film Story. In the story, the child-protagonist steals a basket of pears from an orchard, mounts it on his bicycle, rides along for a while and hits a stone. Together with the basket he falls onto the ground and the pears scatter all over. At the point of the story when (19) is uttered all this has already taken place. Nunga connects the current state of the pears - scattered all over the ground - with the inception of this state, the earlier event of the actual scattering. This use of nunga corresponds to what has been characterised in the literature as "injection" of out-of-sequence events into the main line of the narrative.

(20) is part of the text describing the traditional Batak marriage which, as mentioned earlier, is presented in abstract-generic terms. The two sentences leading to (20) describe a certain stage in the process of marriage, the "mangadati" ceremony. (20) itself diverges from the general mode of presentation in that it does not provide any further information about the sense of a "traditional Batak marriage", but rather makes a comment about the actual present reality with respect to the "mangadati" ceremony. This present reality, (20) says, is characterised by abandoning the ceremony. Nunga relates this reality with an earlier state of affairs when the ceremony was still being performed in the country.

In both examples the earlier situation is presumably known to the audience: in (19), the actual event of scattering the pears is explicitly mentioned previously in the text, and in (20), the fact that the "mangadati" ceremony was performed at one time in Batakland is inferred from the fact that it constitutes a part in the abstract traditional marriage. Thus, like examples (15) and (16), these examples can be viewed as relating in some way the current state to another one known to the audience.

Thus in all examples in this section we have seen predicates depicting a state existing "currently" and nunga relating this state to a previous situation, all in conformity with the definition of the Perfect. Next we will briefly look at another use of nunga, which would still be classified as the Perfect.

1.3 The experiential perfect

When nunga co-occurs with hea (*ever*) in the same clause, the situation expressed by the clause is understood as having occurred "at least once during some time" prior to the time of utterance (cf. Comrie 1976). Thus this type of clause would be used if what is relevant to the discourse is having had the experience of the situation in question. Thus note the contrast between the uses of (21a) and (21b):

(21)a. NUNGA mangallang dengke ho?
eat fish you
Do you already eat fish?

b. NUNGA hea ho mangallang dengke?
Have you ever eaten fish?

(21a) would be licensed by a discourse situation essentially similar to ones we have encountered so far: any discourse, for example, in which the fact is relevant that the addressee eats fish contrary to what has been previously known. For instance, if the addressee has been known to be sick and forbidden by the doctor to eat fish, a friend may utter (21a) to find out to what extent the patient has recovered, or the doctor may utter it on realising the patient has started eating the forbidden food earlier than directed.

(21b) on the other hand, would be used when the discourse concerns issues to which aspects of the experience of having eaten fish are relevant. Such discourse could be a discussion in which the tastes and consistencies of various kinds of foods are compared, or one dealing with the ease or difficulty of consuming certain foods where the fact that one must be "careful with the bones" while eating fish would be relevant.

(22)a. NUNGA laho si Bill tu/sian America.
go PN *to*
Bill's gone to America.

b. NUNGA hea si Bill tu/sian America.
ever
Bill has been to America.

Similarly (22b) would be used if what is relevant to the discourse are general aspects of the experience of having been in America, such as knowing the language or being familiar with attitudes of its people. This contrasts with (22a) which would be used in regard to Bill's "recent" trip to America. It is, of course, hea (*ever*) which effects the difference in meaning between the (a) and (b) cases above, but nunga would typically be present in contexts such as the ones described. Thus we have here another type of function of nunga which can be regarded as an instance of the Perfect. Below, however, we will note a use of nunga which by no means fits the general definition of a Perfect particle as "relating a state of affairs to a prior situation".

2. A NON-PERFECT USE OF NUNGA

Consider the following examples, each of which expresses an *as-yet-unrealised* situation:[12]

(23)a. NUNGA sae (be) ahu.
finish PRT *I*
I'm finishing.

b. NUNGA turun (be) ahu.
come-down
I'm coming.

(24) NUNGA ro be angka tamue.
come PRT PLR *guest*
The guests are coming.

Both sentences in (23) may be uttered in a situation where some person is waiting for the speaker to finish up whatever he is occupied with, and the latter tries to assure the one waiting that he will not take much longer, that he is about to

finish and just about (ready) to come down and join them, so "please, stay calm and do not worry!". There is a sense of immediacy the speaker is trying to convey, the sense that the as-yet-unrealised event is soon to be realised, so soon, in fact, as to be considered happening "now".

Similarly, an adequate context for (24) may be one where expected guests can be seen nearing the house of the host. The anxious host utters (24) in order to urge his daughter or son to take some quick action such as setting up the table, or hurrying up and hiding the dirty clothes somewhere. Again, a feeling of immediacy is noticed.

Another example of nunga in a clause denoting an upcoming situation is found in the Pear Story, mentioned earlier:

(25) ... sae songon i, NUNGA LAHO MA AHU, inna kan.
after like that go PRT *I say*
... after that, "I'm going" they said.
(*Pear Film* 33)

The sense of nunga laho ma ahi here is *let's go* or *let's go now* or *then, let's go*.[13] The act of leaving has not yet taken place, but is, or at least is desired to be, taking place "shortly".

The question, of course, is, what is the function of nunga in this type of context? One way of analysing it - the one I adopted in an earlier version of this paper (cf. Mordechay 1984) - is to view these cases as examples of the "future-oriented" or Prospective Perfect (cf. Comrie 1976), the symmetric analogue in the opposite direction of the more familiar "past-oriented" or Retrospective Perfect. According to this view, the focus of attention of the discourse is, as with the common Perfect, the current state of affairs, but nunga now effects an (inseparable) relation between this state and a subsequent situation.

The other possibility is to analyse these cases as instances of a totally different function of nunga, one which is not temporal (or aspectual) in its essence, but rather, modal. According to this view, nunga would carry a *concessive* meaning, whereby its presence in a clause confirms, assures or admits the factuality of the situation expressed by the rest of the clause.

Both these functions could have evolved from the Perfect meaning through pragmatic extension,[14] and at this point I am not ready to commit myself to either one. But I should point out here that the idea that nunga might have a concessive meaning is based on studies of the German particle *schon* (cf. König 1977, Hoepelman and Rohrer 1981) which, along with a set of temporal-aspectual particles in various other languages - including *already* in English, *déjà* in French, and *kvar* in Hebrew - shows a close affinity with nunga. We will examine now what it is in nunga which makes it similar to these particles.

3. NUNGA AND ALREADY

As mentioned earlier, Percival (1981) translates nunga as *already* and backs his position by citing several examples. Furthermore, looking back at the examples adduced so far in this paper, we see that here too, many of the glosses have included *already*. Moreover, it has been shown that for most of these examples, the function of nunga could be adequately analysed as marking the Perfect, and there has been no sense of conflict with having *already* in the gloss.[15] However, this is not always the case, and in some sentences - in

context - the presence of nunga - and for that matter, the gloss as *already* - would not be appropriate, notwithstanding that the sentence may be interpreted as being in the Perfect. Thus consider the following exchange:

(26) Q: Boasa ndang disurat ho esami?
why not write you your-essay
Why aren't you writing your essay?

A: (*NUNGA) mago pulpenhu.
lost my-pen
*My pen's (*already) lost.*

A: NUNGA husurat be.
I-write
I've already written it.

(26) may take place in a classroom situation, the teacher inquiring as to why a certain student is not participating in the on-going classroom activity of writing an essay. The student may answer nunga husurat be *I've already written it*, mago pulpenhu *My pen's lost*, but not nunga mago pulpenhu *My pen's already lost*. The same holds in the English counterpart of this example replacing nunga with *already*.

Similarly, in the following exchange nunga may only be used in one type of answer but not in the other, and *already* would work analogously in English:

(27) Q: Dohot do ho tu bioskop?
with PRT *you to movie*
Are you coming along to the movie?

A: Daong, adong pamili na ro.
no there-is family that come
No, some relatives have dropped by.

A: Daong, NUNGA hubereng i.
no I-see this
No, I have already seen it.

We see, thus, that a well defined context, such as the ones established by the initial questions in these examples, brings to light a further aspect of nunga (as well as *already*) which restricts the occurrence of the particle to only some of the Perfect sentences. We will see, through the examination of nunga in actual texts, that this characteristic shared by nunga and *already* is indeed a basic one, and the key to understanding the function of nunga and its distribution. This shared function relates to schematic processes and evoking expectations.[16]

4. NUNGA AND SCHEMATIC PROCESSES

I would like to suggest that nunga may occur in a clause just in case the situation expressed by the clause can be viewed as one of the stages in a *schematic process*. By a schematic process I mean the conceptualisation into a single idea of a series of commonly interrelated, temporally ordered situations. So when an expression pertaining to a certain schematic process is found in a discourse, expectations arise, regarding the situation types which constitute that schematic process.[17]

For example, we may think of the schematic process (SP) of writing a letter as involving, among other things, the ideas of cleaning one's desk, getting a pen and some sheets of paper, writing the date on the top of one of those sheets, and so on. So when the phrase 'writing a letter' is encountered, expectations regarding all these specific activity types are evoked.

Nunga, I suggest, indicates that the clause containing it satisfies such expectations.

Thus, if we look back at the exchange in (26), for example, we can understand why one response to the question "why aren't you writing your essay?" is adequate while the other one is not. The inadequate answer, nunga mago pulpenhu *My pen's already lost*, by virtue of containing nunga, presumably fulfils some expectation in a certain SP. A plausible SP here might be one which has to do with owning a pen, a SP whose stages may be becoming the possessor of the pen, having it, and then losing it. But such SP or any other one which includes the idea of losing one's pen has not been evoked in the discourse. Instead, the initial question evokes expectations regarding the process of writing a classroom essay. Thus any sentence containing nunga which pertains to some stage in such a process is adequate, in particular, the one indicating the accomplishment of the goal of the process, i.e. nunga husurat be *I've already written it.* By similar reasoning we can explain the adequacy and inadequacy of the two responses in the exchange cited in (27).

I wish now to look at some occurrences of nunga in actual discourse and to show that, indeed, whenever it occurs in context a recourse to a certain SP is possible.[18]

Two clear examples of this function of nunga are found in (12c) repeated below for convenience. (12c) is in fact part of a narrative which as a whole focuses on a process - that of getting married in Batakland - and explicitly treats this process as a SP. The narrative describes the different stages some members of the community must go through so that the sum of these may be called a "traditional marriage". The two instances of nunga here imply that the two situations in question are to be expected in a traditional marriage and they point to stages where these expectations have been fulfilled.

(12)c. Jala, molo NUNGA rupani denggan-denggan dihatai, jala natua-tua NUNGA
and if as-we-say good discuss and parents

nagihut di hata ni anak, ba adong ma i na ginoarnna manungkun
follow word of son PRT *there-is* PRT *that* COMP *called price*

boru.
woman
And if, as we say, it has been well discussed and the parents already follow the words of the son, then there is the so-called bride-price.
(*Pangolion* 7)

In the same text we find a reference to the inception of pregnancy. Pregnancy is certainly perceived as a process with various phases, universally. Hence the warrant for nunga in:

(28) Dung denggan-pamatang ni boru, jala molo NUNGA marumur pitu bulan
after good-body(pregnant) of daughter and if age seven month

na diboanna i, ba ro ma simatua ni anak on, ...
that she-carry this PRT *come* PRT *relative of son this*

After the daughter becomes pregnant, and if what she carries is already seven months old, the in-laws of the son come, ...
(*Pangolion* 52)

Another piece of text which may easily be conceived as a SP is found in the *Biography* which, as the title suggests, tells the life story of the narrator. In this piece, the narrator describes how he became a "real villager". For someone to deserve the title of a "real villager" he is expected, among other things, to be married. Thus the use of nunga in (29).[19] The same sentence without nunga would indicate equally well that the narrator almost got married, but without implying that this is an expected thing to do in the process of becoming a "real villager".

(29) Martani ma au disi sataon. Jala au ma disi tungga ni huta ..., boha ma
farming PRT *I there one-year and I head of village how*

a NUNGA gabe parhuta-huta attong iba kan? Jadi, sapotulna NUNGA
become villager PRT *self you-know so actually*

happir[20] mangoli au.
almost marry I
I was farming there for one year and I was the head of the village .... Do you know I was already becoming a real villager. In fact, I almost got married.
(*Biography* 19-24)

The next example is taken from a text titled *My plans for the future* and is perhaps a slightly less explicit case of a SP. The narrator expresses his desire to become a teacher, and describes his preparation, and the development of his affinity for a teaching career. The childhood playing of the role of a teacher is viewed as part of this development:

(30) Alana sian na jolo minathu na sai bagas, ima mangajar. Ai
because from that early my-desire that so big it-is teach PRT

boha ma natua-tuaniba guru do. Jala tingki metmet iba NUNGA marsiajar
how PRT *my-father teacher* PRT *and when small self study*

mangurui marhite hagioton-hagioton di pagarejaon.
teach through activities in church-activities
Because since early on I had the desire to be a teacher. My father was a teacher, and when I was small I was already studying how to play the role of a teacher through participating in church activities.
(*Plans* 7-9)

Another example which deals with some aspects of the process of bringing up a child is the one in (31). The narrator comments on the fact that the children in the depicted world return some pears to the believed owner instead of keeping them for themselves as he would expect. Thus he commends the process of disciplining which is responsible for such honesty at such an early age.[21]

(31) Bah, heran ... aha kehidupan ni dakdanak on songon ..., gabe
PRT *surprised*(I) *what life*(I) *of child this like become*

heran do iba, arti-na dakdanak i NUNGA ... diajari ... pittor.
surprised(I) PRT *self mean child the study honest*
I was surprised ... what the child's life is ... I was surprised, I mean the child is already taught to be honest.
(*Pear Film* 33-35)

The next example is from the text *Daily life in Samosir* which, as its name suggests, describes the everyday life routine in the island of Samosir. This is again a rather explicit description of a SP, and in this particular one the idea of getting up in the morning is the most obvious stage. Thus:

(32) Ai jam lima monogot ba NUNGA antong dungo iba.
PRT *hour five morning* PRT PRT *wake-up self*
At five o'clock in the morning, I'm already up.
(*Daily life in Samosir* 3)

In a similar fashion a SP may be reconstructed for each of the occurrences of nunga in the attested discourses.[22] Thus I take the notion of a SP as fundamental to understanding the function of nunga. This notion would further explain an additional use of nunga.

When nunga is stressed in a sentence, the sentence receives a sense of comparison so that its translation into English would include expressions like 'as soon as', 'as early as', 'as far as', etc. Thus if we stress nunga in (31) for example, we get a sense something like "already at that early age, children are taught to be honest". Such sense arises when a given situation is compared with one which is expected according to some relevant SP. A SP, as said, is the conceptualisation of temporally ordered situations, so speakers have expectations as to which situations come commonly before others, and which come later, and in general have a sense of what is appropriately regarded as 'early' or 'late'.

If the situation described in the clause is supposedly at the expected time or at an earlier one, the use of nunga is appropriate. But if it is later than the expected time, nunga may not be used. Thus if we change jam lima *five o'clock* in (33) to jam san pulu sada *11 o'clock*, the sentence would be inappropriate for most contexts.

(33) NUNGA jam lima, dunggo iba.
hour five wake-up self
Already (as early as) at five o'clock, I'm up.

(33) demonstrates an additional property of nunga, namely, its ability to combine not only with clauses but with time adverbials as well. As we will see in the next few final examples, nunga may, in fact, combine with phrases of various types and yield meanings which have nothing at all to do with time.

Consider the following:

(34) Holan manuhor orbuk pe halak, ai NUNGA ikkon manggarar rubuana rupia.
just buy fertiliser people must pay thousands rupees
Even (in order) to buy fertiliser, one must pay now as much as thousands of rupees.
(*Conditions in Indonesia* 23)

(34) is taken from the text mentioned earlier which described the severe conditions in Indonesia since the generals took over power. This sentence has two readings, both of which are compatible with the surrounding context and with each other. The immediately preceding sentence says that the conditions of agriculture in the rural areas have become difficult. (34) can, on the one hand, be interpreted as implying a process of economic deterioration: "The economic deterioration has reached that point now where even fertiliser costs thousands of rupees". This interpretation involves a time axis.

But there is another interpretation which does not involve time, and whereby the sentence may be viewed as a comment on the economic conditions from a synchronic point of view. Under this interpretation, a given price of a product is compared to a preconceived assessment of that product. Such assessment is construed as the different products in the market are conceptually ordered along a price scale, and thus speakers have an idea of which products are 'cheap' and which are 'expensive'. According to this interpretation, (34) means something like: "Even something as cheap a product as fertiliser costs as much as thousands of rupees."

This example might not be convincing as an instance of the *non-temporal* use of nunga as it includes holan *even* and since it is possible to interpret it in a temporal sense. Thus let us look at (35) where such interpretation is inadequate:

(35) Molo NUNGA sahat 500, boasa ndang 530 sahali, asa dapot sude?
if arrive why not at-once so-that get all
If (you are willing to pay) as much as 500, why not (i.e. you might as well) pay 530 and get everything?

The context here might be a shopping situation where some customer is hesitating as to how much he is willing to spend. He finally decides on a relatively high price but not high enough to get him what he really wants, so a friend utters (35). Nunga may be used here only in the case that 500 rupees are indeed preconceived as a large amount of money for the customer in question.

Similarly, the two examples in (36) involve comparisons along some pre-established axes which are not time axes.

(36)a. Molo NUNGA sahat di New-York, boasa ndang laho tu Long-Island?
if arrive in why not go to
If we (intend) to get as far as New York, then we might as well go to Long Island!

b. Molo NUNGA si Reagan, boasa ndang si Jerry Falwell sahali?
If (you're going to vote for someone as far to the right as) Reagan, then why not (go the whole way and vote for) Jerry Falwell?

(36a) points to a conceptualised ordering of places by the distance from the discussants, an ordering according to which both New York and Long Island are regarded as far away, but Long Island is further. (36b) implies a preconceived ordering of persons according to their political inclination, or better, their inclination toward the political right. In each of these cases, nunga is appropriate only if the idea of the person or place in question is one of the constituents which make up the pre-established ordering.

Thus in analogy to the notion of a schematic process, we may introduce the more general notion of a *schematic ordering* of which a SP would be just a special case. We may say, then, that a major function of nunga is to indicate that the phrase (or clause) with which it combines denotes something whose sense is one of the constituents comprising a schematic ordering.

5. SUMMARY

I have attempted to shed some light on the working of the aspectual particle nunga in Toba Batak. I hope this purpose has been achieved through the description of the contexts and discourses in which the particle is, or may be, used.

Two analyses of the function of nunga have been proposed. According to one, nunga is a Perfect marker, and according to the other, it implies a schematic ordering. The latter characterisation is both more general and less general than the former one. It is more general since it is not restricted to the temporal use of nunga or to a relation between two situations only, two properties characterising the Perfect. But it is also more specific since for the temporal use, it focuses on a particular property - the one regarding schematisation - which is not a common property of the Perfect. However, as suggested, this property is essential in our understanding of nunga.

NOTES

* I wish to acknowledge a special debt to Paul Schachter for his stimulation and guidance during the writing of this paper, and to Wilson Manik who provided the data reported on here with patience and insight. Thanks are also due to Jack DuBois, Ed Keenan, Christian Matthiessen, John Singler, Sandy Thompson and Alan Timberlake for useful discussions and comments.

1. See, for example, Anderson 1972, Hoepelman and Rohrer 1981, König 1977 and Li, Thompson and Thompson 1982.

2. Wilson Manik was the source of both elicited data and recorded texts. He also helped glossing and interpreting the latter texts.

3. The brief description of the Perfect category which follows is my understanding of what is meant by Perfect in the literature on aspect, and is based primarily on Comrie 1976 and McCoard 1976, as well as on Anderson 1982, Givón 1982 and Li, Thompson and Thompson 1982. While such description would probably be classified as belonging to the "Current Relevance" theory of the Perfect, I have intentionally avoided the term "relevance" accepting the criticism made in McCoard 1976 that it does not further our understanding of the Perfect.

4. Although (1d) does not have the form of the English Perfect, it has the Perfect meaning; and in many languages such sentences do appear in the Perfect form. See below for discussion of this example.

5. As is common now in the linguistic literature on aspect, following Vendler's (1967) distinction of verb types, I adopt a terminology where *situation* is used as the neutral cover term to include states, activities and events. (And see, further, for example, Comrie 1976 and Mourelatos 1981.) *States* are static non-changing situations. *Activities* are dynamic ongoing situations, and *events* are dynamic situations viewed in their entirety, as an individuated something. Usually in the literature of aspect, the terms *process* and *activity* are interchangeable. However, in this paper I maintain the distinction between them as in their conventional, non-linguistic use, where *process* suggests directionality but *activity* does not. I use *activity* for the verb type as defined above, and reserve *process* for a later use.

6. See Reichenbach 1947, who introduced the term *reference time* as distinct from speech time and event time in order to distinguish between the Perfect and the non-Perfect tenses.

7. The following abbreviations are used in glosses: PN = Proper Noun Marker; PRT = Particle; PLR = Plural; COMP = Complementiser.

8. Thanks to J.P. Sarumpaet who pointed this out to me during the presentation of this paper at FOCAL. The predicate particle ma functions as a sequentialiser and may be glossed as *then*. See further discussion on ma in Jackson 1984 and Mordechay (in preparation).

9. Nunga is occasionally accompanied by the particle be which occurs at various positions in the clause. Percival (1981) glosses be as *already*, but in the current work its distribution and function are not further studied.

10. A statistical correlation exists between the aspectual categories of the perfective and imperfective and the two major transitive verb forms, the di- (patient oriented) form, and the mang- (actor oriented) form. However, aspectual distinction is probably not the only major function of the verbal morphology distinction. In fact, Wouk (1984) has argued to the contrary that the property of patient individuation is the major factor responsible for the distinct verb forms. But Wouk grouped together verbs in the di- form with those in the ni- form. The ni- form, however, is a nominal form and has a stative, i.e. imperfective meaning. Thus this grouping skews the results against correlation with aspect. (cf. Tuller 1984 for a discussion on the syntax of ni- verbs.)

11. Toba Batak has a variety of verb forms beside the aforementioned mang, di and ni forms, some of which would be considered as having an imperfective meaning, and others, perfective. However, their frequency is much lower than that of the former three. See e.g. Percival 1981 for a description of the other forms.

12. In many cases, including the ones at hand, it is impossible to determine the tense of a sentence, when looking at it in isolation, since tense is for the most part not marked in TB. As usual, context solves the indetermination.

13. The particle ma occurs often in imperatives. See also note 8.

14. Alan Timberlake drew my attention to such a possibility for the case of the concessive meaning.

15. Indeed in English, when *already* is present in a sentence, the sentence is usually in the Perfect, as in e.g. *I have already read this book.*

16. The analysis of nunga which follows is similar to the analysis of the German particle *schon* in the works of König (1977) and Hoepelman and Rohrer (1981), though our terminology is different.

17. I use "schematic process" in roughly the same sense of "script", "frame", or "schema" now used in Discourse Analysis. See for example, Chafe 1984 and Schank and Abelson 1977.

18. My data include 16 cases of nunga, all except one (see note 22) attest this function of the particle. While this is certainly not a large sample, further examination of data in hypothetical contexts verifies the observation.

19. I refer here to the second use of nunga in the paragraph. The first one pertains to a less explicit schematic process, namely, one which is construed by the expectation that the narrator will become a real villager as a result of knowing that he has spent a full year farming in a village.

20. Happir is Indonesian for *almost*.

21. The text here is fragmented and some of the words are Indonesian. They are marked in the gloss by (I).

22. Excluding one case of nunga analysed either as the concessive function or as the "future-oriented" Perfect.

REFERENCES

ANDERSON, Lloyd

1982 The 'perfect' as a universal and as language-particular category. In Hopper, ed. 1982:227-264.

ANDERSON, S.R.

1972 How to get even. *Language* 48:893-906.

CHAFE, Wallace

1984 *Cognitive constraints on information flow*. Berkeley Cognitive Science Report No.26.

COMRIE, Bernard

1976 *Aspect*. Cambridge: Cambridge University Press.

GIVÓN, Talmy

1982 Tense-aspect-modality: the creole proto-type and beyond. In Hopper, ed. 1982:115-163.

HOEPELMAN, J. and C. ROHRER

1981 Remarks on noch and schon in German. In Tedeschi and Zaenen, eds 1981:103-126.

HOPPER, Paul, ed.

1982 *Tense-aspect: between semantics and pragmatics*. Amsterdam/ Philadelphia: John Benjamins.

JACKSON, Catherine

1984 Focus particles in Toba Batak. In Schachter, ed. 1984:80-99.

KÖNIG, E.

1977 Temporal and non-temporal uses of 'noch' and 'schon' in German. *Linguistics and Philosophy* 1:171-198.

LI, C., S. THOMPSON and R. THOMPSON

1982 The discourse motivation for the perfect aspect: the Mandarin particle *le*. In Hopper, ed. 1982:19-44.

McCOARD, Robert

1976 Tense choice and pragmatic inference: a study of preterit/perfect opposition in English. Ph.D. dissertation, UCLA.

MORDECHAY, Susan

1984 The semantics and pragmatics of the "perfect" marker *nunga* in Toba Batak. In Schachter, ed. 1984:100-121.

in preparation Predicate particles and narrative structure in Toba Batak.

MOURELATOS, Alexander

1981 Events, processes, and states. In Tedeschi and Zaenen, eds 1981: 191-212.

NABABAN, P.W.J.

1966 Toba Batak, a grammatical description. Ph.D. dissertation, Cornell University.

1981 *A grammar of Toba-Batak*. *PL*, D-37.

PERCIVAL, W.K.

1964 A grammar of Toba Batak. Ph.D. dissertation. Yale University.

1981 *A grammar of the urbanised Toba-Batak of Medan*. *PL*, B-76.

REICHENBACH, Hans

1947 *Elements of symbolic logic*. New York: Free Press.

SCHACHTER, Paul, ed.

1984 *Studies in the structure of Toba Batak*. UCLA Occasional Papers in Linguistics, No.5.

SCHANK, R. and R. ABELSON

1977 *Scripts, plans, goals, and understanding*. Hillside, N.J.: Lawrence Erlbaum.

TEDESCHI, P. and A. ZAENEN, eds

1981 *Syntax and semantics*, vol.14: *Tense and aspect*. New York: Academic Press.

TULLER, Laurice

1984 Genitive case in Toba Batak. In Schachter, ed. 1984:172-194.

VENDLER, Zeno

1967 *Linguistics in philosophy*. Ithaca: Cornell University Press.

WOUK, Fay

1984 Scalar transitivity and trigger choice in Toba Batak. In Schachter, ed. 1984:195-219.

Mordechay, S. "The aspectual particle nunga in Toba Batak". In Geraghty, P., Carrington, L. and Wurm, S.A. editors, *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*. **C-93:113-132**. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.113*

TRANSITIVITY IN PROTO-MALAYO-POLYNESIAN AND PROTO-AUSTRONESIAN

Fay Wouk

INTRODUCTION[1]

Many Austronesian languages are characterised by opposition between clauses with differing verb morphology, often called a focus system. In this paper I look at how this opposition of clause types functions in each of a variety of Austronesian languages, to determine the functional common ground between Oceanic and Western clause types, and suggest the function of the earlier system in Proto-Austronesian (PAN).

Reconstructions of the PAN "focus system" are almost as common as subgrouping assumptions, and equally varied. The goal of such reconstruction has been to find the structural and morphological common ground between Oceanic and Western clause types, and to establish the structure and morphemes of PAN clauses. There is a general consensus about what the common ground is, although there is considerable disagreement as to its correct interpretation. My research into the functional aspect supplements this effort.

The Oceanic clause type is characterised by the use of pre-verbal subject cross-referencing, and with transitive verbs by the use of post-verbal object cross-referencing[2] and two verbal suffixes, reflexes of *-i and *-aken.[3] The choice between the suffixes depends on the semantic role of the object. *-i is found when the object is patient, goal or location. *-aken is found when the object is an instrument or beneficiary.

The Western clause type (found in many Indonesian, Philippine and Formosan languages) is not characterised by cross-referencing of any kind, although it may occur. The most noticeable characteristic of the Western clause type is the "focus system", in which the "subject" may belong to one of two or more sets of semantic roles, and a verbal affix identifies which set of semantic roles the "subject" belongs to. The affixes used in the Western clause type are reflexes of *-um-, *-in-, *-en, *-an and *Si-. *-um- indicates agent or experiencer. *-in-, earlier a perfective marker, now indicates patient in many languages. *-en indicates patient in some Philippine languages. *-an indicates location or goal. *Si- indicates instrument or beneficiary. Some Western languages (generally outside the Philippines) also use reflexes of *-i and *-aken to further specify the semantic role of the underlying object of the verb. They may be used in conjunction with reflexes of both *-um- and *-in-.

There are four major hypotheses about the structure of PAN clause types. These hypotheses are outlined below. (The names I have given to the hypotheses

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 133-158. *Pacific Linguistics*, C-93, 1986.

should be considered as no more than mnemonics. I do not claim, nor did the authors of the papers cited, that the structure of any modern language or language group is identical to PAN.)

Oceanic hypothesis: Foley (1976) proposes a reconstruction according to which PAN used only the suffixes *-i and *-aken. As in modern Oceanic languages, the choice of suffix depended on the semantic role of the object NP. With one set of semantic roles *-i would be used, while with another set *-aken would be used.

Philippine hypothesis: Wolff (1979) proposes something more akin to the Philippine system, according to which PAN had a four-way focus system using the affixes *-um-, *-in-, *-an, and *Si- to give information about the semantic role of the focused NP. As in the Oceanic hypothesis, each affix is associated with a set of semantic roles.

Toba Batak hypothesis: Pawley and Reid (1979) propose what may be called the Toba Batak system, since they suggest that Toba Batak as described by van der Tuuk (1971) is the modern language which bears the greatest resemblance to their reconstruction. This system makes use of *-um- to indicate actor focus, and uses *-i and *-aken in actor focus to give information about the semantic role of the object NP. In the non-actor focuses *-in-, *-an and *Si- give information about the semantic role of the focused NP.

Formosan hypothesis: Starosta, Pawley and Reid (1982) propose a hypothesis in which *-i and *-aken were the PAN focus markers, each one associated with a set of semantic roles, while *-um-, *-in-, *-an and *Si- were nominalisers which were later reinterpreted as focus markers in a number of daughter languages through a process of drift.

The common ground that all these hypotheses find between the two modern systems, Eastern and Western, is the similarity between the sets of semantic roles associated with certain morphemes. Specifically, *-i in Oceanic languages is associated with the same set of semantic roles as *-in- and *-an in Western languages, and *-aken in Oceanic languages is associated with the same set of semantic roles as *Si- in Western languages.

With all the reconstruction attempts, and all the discussion of syntax of PAN that has gone on, it is nonetheless rare to find a discourse/functional analysis of either the Western type or the Eastern type, or any discussion of functional similarities between the two. The only exception I am aware of is Harvey (1982). He proposes that since PAN was verb initial, this created a situation in which new information (the verb) was in an inappropriate place, namely clause-initial position, which should be reserved for some element providing cohesion with the previous clause. According to his analysis the focus affixes provide this cohesion, since they refer to an element which is typically old information. He claims that "from the semantics of the affix and the case frame of the verb one is able to tell which NP will be the PrP [pragmatic peak] and so one knows which pieces of information from the previous clause(s) are the centre of this new clause" (p.67).

Neither Harvey nor anyone else questions the assumption that the Western system is a "focus" system, with a function of indicating that NP which is in some (unspecified) sense the most important one in the clause. Unless this assumption is correct, Harvey's analysis cannot hold. In this paper I will show that this assumption is not valid, and thus that a different function must be postulated for verbal focus. I will present a different analysis of the common ground between the two systems, and of the function of both of them.

My analysis is discourse based, and stems from Hopper and Thompson's (1980) theory of transitivity. I will attempt to show that the two systems share not only a similar subdivision of case roles into grammatical categories, but also a similar discourse function for their overall clause structure, having to do with the expression of discourse transitivity. Many Austronesian languages characterise and express discourse transitivity in a similar way, which is peculiar to the Austronesian family. The most important parameter appears to be individuation of the patient, not presence or absence of a patient. High transitivity is marked morphologically on the verb, using one of the two sets of affixes described above. I will also show that this function can be reconstructed for Proto-Malayo-Polynesian (ancestor of modern non-Formosan languages), and will suggest that it can also be considered for Proto-Austronesian (ancestor of all Austronesian languages, Formosan and extra-Formosan).

TERMINOLOGY

Before going on with a discussion of the issues, I will comment on some of the terminology that I will and will not be using. Since my analysis stems from a discourse/functional approach to language, rather than from the orientation of some formally-based syntactic theory, I will use terms with functional definitions. Thus an NP is not a subject unless it can be shown that it has the same function as subjects in other languages in which the term has accepted use, such as English. In the three Austronesian languages for which I have done discourse studies (Tagalog, Toba Batak, and Malay-Indonesian), the "focused" NP does not have the same function as the subject in English, and object focus does not have the same function as passive. English subjects (Givón 1979) are thematic, and show a high degree of continuity throughout a text, which is not true of English objects as a rule. Research by Thompson (forthcoming) shows that the use of English passive "seems to be based on the discourse structuring principles of thematic and inter-clausal continuity" (p.19 of draft copy). Passive subjects tend to be thematic, while passive agents are non-thematic, and the function of passive is to keep thematic elements in the subject slot. However, "focused" NPs as a class in the three languages studied were not significantly more thematic than were agents as a class, and in fact the "focused" patient NP was often less thematic than the non-focused agent.

In Indonesian (Wouk, forthcoming b) patient focus is obligatory in an independent clause when the agent is not present. Yet in many cases the agent is deleted because it is the most topical, most thematic element in the clause. In Batak and Tagalog (Wouk, forthcoming a) patient focus correlates with individuation of the patient, and thus is used even when the agent is highly thematic. In Tagalog (Cooreman et al. 1984) agents have been shown to be more topical than patients according to continuity measurements. One of the measurements used is lookback, which can be defined as the number of clauses between a given mention of an NP and the previous mention. NPs with shorter lookback are more topical than those with longer lookback. Among clauses with verb-initial order, agents had an average lookback of 2.88 clauses in patient focus, and 1.62 clauses in actor focus. Patients had an average lookback of 10.10 in patient focus, and 19.01 in actor focus. Thus, while patients in patient focus are more continuous (i.e. topical) than those in actor focus, all patients are much less continuous (i.e. topical) than all agents, regardless of verbal focus.

Because of the functional implications mentioned above, I will not use the term subject to refer to the "focused" NP. Nor will I use the term topic, which

has similar implications. "Focused" NP is also problematic, since focus is generally used for new information, but it is not the case that "focused" NPs are predominantly new information. "Focused" NPs are generally old information. The term I will use is "trigger", which was suggested by Fox (1982) on the grounds that it is the NP which "triggers" verb morphology.

There has also been much discussion in Austronesian studies in recent years of ergativity (Hohepa 1969, Clark 1976, Chung 1981, Cooreman et al. 1984, Verhaar forthcoming) with suggestions that various modern languages or earlier stages of languages should be considered ergative. While this may or may not be true and relevant for the synchronic description of individual languages, I feel that in the context of this paper this is not a substantive issue. The substantive issue is: what is the function of verbal morphology in a given Austronesian language, and is it the same as in other Austronesian languages, or different? This is the issue I wish to address. In terms of the system I reconstruct, there is no significant difference between "ergative" languages like Tagalog and "accusative" languages like the Micronesian ones. They all distinguish clauses on the basis of level of transitivity. In fact, there is more difference between the "ergative" Polynesian languages, which show only a relic transitivity distinction rather than a productive one, and "ergative" Tagalog, than difference between "ergative" Tagalog and the "accusative" Micronesian languages. I will therefore ignore the issue of ergativity.

DISCOURSE TRANSITIVITY

According to Hopper and Thompson (1980), transitivity is scalar. The level of transitivity of a given clause is not just a result of the number of arguments that a verb has, but rather of a combination of factors, or parameters, all concerned with the effectiveness with which an action is transferred from an agent to a patient, such that verbs with the same number of arguments could be higher or lower in transitivity with respect to each other, depending on the status of the other parameters. Hopper and Thompson found that these parameters tend to co-vary, so that a clause that is high in one will usually be high in several others. A given clause is considered more transitive if it ranks high on a number of these parameters, and less transitive if it ranks low on many of them. The parameters that they consider relevant are given in the following table.

Table 1

	HIGH	LOW
participants	two or more	one
kinesis	action	non-action
aspect	telic	atelic
punctuality	punctual	durative
volitionality	volitional	non-volitional
affirmation	affirmative	negative
mode	realis	irrealis
agency	high in potency	low in potency
affectedness of patient	totally affected	unaffected
individuation of patient	highly individuated	non-individuated

Highly individuated patients are opposed to less individuated patients according to the following parameters:

Table 2

INDIVIDUATED	UNINDIVIDUATED
proper	common
human, animate	inanimate
concrete	abstract
singular	plural
count	mass
referential, definite	non-referential

Hopper and Thompson also show that crosslinguistically, morphosyntax is often sensitive to the overall level of transitivity of a clause, rather than just to the number of arguments present. Morphological marking in many languages differentiates "transitive" and "intransitive" on bases other than number of arguments. This makes it desirable to refer to high transitivity clauses and low transitivity clauses, rather than the traditional transitive and intransitive. Transitivity marking may involve verbal morphology, case marking, nominal particles, special sets of pronouns, word order, or perhaps other strategies. A distinction may be made in past or perfect two-argument clauses only, or two-argument clauses with punctual verbs only, or only in clauses involving kinesis and affectedness of patient, or in clauses with a definite, or referential, or animate patient only. Thus, while there are universal correlations between transitivity marking and a wide variety of parameters, there is considerable variation in the detail of transitivity marking. It is quite possible for two languages to have distinctive patterns of transitivity marking, even though neither of them defines transitivity in terms of simply having two arguments. A high degree of similarity both in terms of the type of transitivity marking and in the particular parameters that are emphasised is unlikely to occur coincidentally.

TRANSITIVITY IN MODERN AUSTRONESIAN LANGUAGES

In this section I will discuss the expression of transitivity in a number of modern languages and language groups. I will attempt to show both that in many Austronesian languages it is discourse transitivity that is relevant, not just the number of arguments in the clause, and that the parameters and types of marking are highly similar throughout the family. The groups in question reflect geographical divisions, and do not necessarily imply genetic subgroups, although they do not preclude them.

Before I begin, I would like to comment on the data presented in this paper. Some readers may be puzzled that the example sentences do not always show perfect parallelism in the choice of lexical items. This is because most example sentences are taken from a variety of reference grammars and articles, while those for Batak and Tagalog are taken from texts produced by native speakers. The examples in question were not necessarily originally intended to illustrate the points discussed in this paper. For this reason, examples are not always perfectly parallel. The variation in choice of lexical items has no particular significance beyond the chance inclusion or exclusion of a sentence in my sources.

Also, with a few exceptions, most of the following discussion is not accompanied by actual text counts. Much analysis in the past has been based only on isolated sentences, and on general impressions, rather than on careful examination of texts. It thus presents only a partial picture of the languages involved. Ideally, this should be supplemented with textual data.

PHILIPPINE LANGUAGES

At least two Philippine languages, Tagalog and Ilokano, show a correlation between trigger choice and patient individuation. In Ilokano, according to Constantino (1971), the trigger must always be definite. A non-trigger patient will be indefinite, while a non-trigger actor or oblique may be either indefinite or definite.

Ilokano

G-um-atang iti mangga diay ubing
AT:*buy*[4] art *mango* art *child*
The child will buy a mango.

Basa-en na ta libro
PT:*read he* art *book*
He will read that book.

It has been held that in Tagalog, a trigger must be definite, and a non-trigger patient must be indefinite, the only exceptions being relative clauses on the actor, which permit definite patients (Schachter and Otanes 1972). However, Bloomfield's earlier analysis (1917) did not draw so sharp a line, and he noted that sometimes patients which might be considered indefinite were triggers. My own research (Wouk 1984) confirms Bloomfield's observation that triggers need not be definite, and also shows that a non-trigger patient occasionally can be. I did a discourse study, using data from Bloomfield's collected texts and from modern written and spoken Tagalog texts. I found that there was a very strong correlation between definiteness of patient and patient trigger morphology, and between no patient and actor trigger. However, when the patient was indefinite, there was a certain amount of indeterminacy. The tendency was for non-referential patients to be found in actor trigger clauses, while specific indefinite patients were evenly split between actor trigger and patient trigger.

Referential and non-referential are used here and throughout the paper to refer to discourse referentiality, as defined by Dubois (1980), not semantic referentiality. An entity has discourse referentiality if within the text it is treated as referential, allowing it to be replaced by anaphora after its initial mention. Something can be semantically referential, and yet not be so treated in a text.

Tagalog

T-in-awag nila ang isang kalaw upang siya ng maging hukom
PT:*call they* art *one owl so he* art *become judge*
They called an owl, so he could be the judge.
(PT with specific-indefinite patient)

kung umaga ay[5] i-ni-hahanda ang kanyang panghilamos
if morning prt PT:*give* art *his washwater*
In the morning he would give him his washwater.
(PT with non-referential patient)

kanya t-in-awag ni Andres ang unggo
so PT:*call* art name art *monkey*
So Andres called the monkey.
(PT with definite patient)

nag-tayo siya ng poso ng tubig
AT:*build he* art *well* art *water*
He built a well.
(AT with specific-indefinite patient)

nag-padala siya nito sa Pilipinas
AT:*send he it to* name
He sent it to the Philippines.
(AT with definite patient)

While Tagalog does not show a one-to-one correspondence between definiteness and patient trigger in discourse, the correlation is quite high, and mathematical analysis showed it to be statistically significant. It is clear that there is a correlation in Tagalog between patient trigger morphology and high transitivity in terms of patient individuation; however, there is some indeterminacy as to the exact cut-off point between an individuated patient and an unindividuated patient which indicates that this is a discourse phenomenon, not a syntactic one. The following table shows the number of clauses with each type of patient for both actor-trigger and patient-trigger clauses.

Table 3: Patient status in Tagalog

	No patient	Non-Ref.	Specific-Indef.	Definite
AT	18	23	14	5
PT	0	13	14	99

Thus, at least two languages of the Philippines show a correlation between verb morphology and individuation of patient. This is similar to the situation in the Oceanic languages, as I will show, although completely different morphemes are involved. The Oceanic system has long been recognised as a transitivity marking system, although the importance of a discourse definition of transitivity has not been universally recognised. However, until recently it has not been suggested that the Philippine trigger marking systems expressed transitivity. I believe that the evidence makes this a necessary conclusion. In the Philippine languages, just as in the Oceanic ones, two-argument transitivity is not indicated, but high transitivity in terms of an individuated patient is morphologically marked on the verb.

INDONESIAN LANGUAGES

Among Indonesian languages, *-i and *-aken function as transitive suffixes, increasing the valence of the verbs they are affixed to, *-i generally having a locational sense, and *-aken a benefactive or instrumental sense. These suffixes are found as -i and -ake/-aken in Javanese (Poedjosoedarmo 1976), -i and -akan

in Indonesian (Wolff 1980), -i and -aka in Wolio (Anceaux 1952), and -i and -hon in Toba Batak (van der Tuuk 1971). They are by no means required to form transitive verbs, and do not correlate in any way with the trigger marking systems found in any of these languages.

Javanese

Pak guru m-aring-i Tono buku kuwi
father teacher AT:*give* name *book this*
The teacher gave Tono the book.

Tono di-paring-i buku dening Pak Guru
name PT:*give* *book by* *father teacher*
The teacher gave Tono the book.

Batak

man-angis imana
AT:*cry* *he*
He cried.

man-angis-hon hamamate ni innatta i imana
AT:*cry* *death* *of mother this he*
He mourned the death of his mother.

man-angis-i innatta i imana
AT:*cry* *mother this he*
He mourned his mother.

Many Indonesian languages make a morphological distinction between verbs with individuated patients and those with unindividuated patients. Modern Javanese (Horne 1961), Old Javanese (Hopper 1979a) and Early Modern Malay (Hopper 1979b) have all been included in this group, and in the case of Hopper's work, the analysis is based on textual studies. In all three languages, patient trigger morphology correlates with individuated patients, and actor trigger morphology correlates with less individuated patients.

Old Javanese

ma-manggih ta sira yuyu ri sikkhara-ning parwata
AT:*find* emph *he* *crab at summit:of* *mountain*
He found a crab at the top of the mountain.

S-in-ambut ikang yuyu de sang brahmana
PT:*take* *the* *crab* agt *the* *brahmin*
The brahmin took the crab.

Malay

si-Manap ng-urup-kan ringgit pada si-Amin.
name AT:*change* *dollar with* name
Manap changed a dollar with Amin.

Olahnya baru tadi di-pentin de si-Amin-nya ringgit itu
apparently just now PT:*test* *by* name *dollar that*
Apparently Amin just tested the dollar.

Discourse studies of Toba Batak (Wouk 1984) show a similar situation. In this study I found that there was a statistically significant correlation between

patient trigger and both referential patients and individuated (modified) patients. For Batak I found that individuation was a better predictor of trigger choice than referentiality, because there were a number of texts about Batak customs, where the NP could not be said to refer to individuals in the real world, and which I counted as non-referential for this reason. If I had counted these as referential, on the grounds that they existed within the world of the text, that would undoubtedly have made the results nearly parallel to those in Tagalog, where all the texts I used referred to the real world.

Batak

inong niba jolo di-unung do[6] martolu-tolu borngin
mother self before PT:*persuade* prt *three nights*
It took three nights to persuade my mother.
(PT with individuated patient)

sae dungo pa-dungkarhon horbo, sae pa-dungkarhon horbo,
after rise AT:*cause:go:out water:buffalo after* AT:*cause:go:out w:b*

tu juma man-inggala hauma
to field AT:*cultivate field*
After getting up, (I would) let out the water buffalo, after letting out the water buffalo, to the field, (I would) cultivate the field.
(AT with unindividuated patients)

In Batak, as in Tagalog, in actual texts there was not a perfect correlation between individuation of patient and trigger choice. This indicates that in both languages this is a discourse phenomenon, not a syntactic one. High transitivity marking is being used for a discourse function, indicating the salience of the patient to the discourse. Salience is defined as the degree of prominence of a participant in a discourse. Table 4 gives figures for patient type in Batak with AT and PT clauses classified in terms of patient individuation.

Table 4: Patient status in Batak

	no patient	unindividuated	individuated
AT	24	34	13
PT	6	5	136

It should be noted that it is not possible, in either language, to correlate patient trigger morphology with greater salience of the patient than the actor. Textual analysis shows quite clearly that actors are more thematic and continuous than patients, even when the patient is the trigger. It is absolute salience of the patient that is at issue here, just as in Oceanic transitive marking. The morphology registers the fact that the speaker deems this participant to be prominent, to be figure rather than ground in the text that speaker is producing.

RELATIVISATION IN INDONESIAN AND PHILIPPINE LANGUAGES

In many Western languages with trigger marking systems the head of a relative clause must be the trigger of that clause. In some Philippine and Indonesian languages in patient trigger the patient, but not the actor, may be head of the clause, and in actor trigger the actor, but not the patient, may be. Among the Philippine and Indonesian languages, this has often been taken as evidence that

the trigger is the subject, since it is the only direct argument NP which is accessible to relativisation. Schachter (1976) has pointed out problems with this analysis, since there are other processes to which only the actor is accessible. I believe that there is a simple explanation to this particular anomaly, which relates to transitivity.

Any NP coreferential with the head of a relative clause is by definition individuated, since its identity is established prior to the relative clause, and the relative clause itself further specifies the identity of the referent. (It is not the case, however, that the head of the relative clause, whether coreferential with actor or patient within the clause, need itself be referential.) Thus, if the NP which is head of a relative clause is coreferential with the patient in that clause, the verb must be marked for high transitivity, because its patient is individuated. On the other hand, if the head of a relative clause is coreferential with the actor, the patient may either be individuated or not, thus allowing two options for verb morphology. There is a conflict here between accurate marking of patient status and degree of transitivity, and maintaining a maximally simple relativisation strategy. I suggest that at some point in the history of PMP this conflict was resolved in favour of the simple relativisation strategy. Many daughter languages retain this constraint, or a trace of it, regardless of whether or not they have clearly identifiable subjects.

MICRONESIAN LANGUAGES

Most Micronesian languages have undergone considerable phonological change, such that PAN morphology is not readily recognisable. However, most potentially two-argument verbs have two forms, generally called transitive and intransitive. These classes often show phonological alternations or have suppletive forms which reflect the earlier presence of *-i and *-aken on the "transitive" forms. In some languages suffixes consisting of a thematic consonant and /i/ are added to some transitive verbs. This suggests that the "transitive" and "intransitive" verbs can be derived from earlier forms with productive suffixation. The most striking thing about all these languages is the correlation between transitive verb forms and individuated patients.

One group of languages, including Marshallese (Bender 1969) and Trukese (Sugita 1973) follows a pattern whereby those verbs identified as intransitive appear without a patient, with an unindividuated patient, or with an individuated patient when the sense of the clause is partitive. "Transitive" verbs appear only with specified patients, and the patient must always be totally affected.

Trukese

Wupwe mwenge rayis
I:will eat(i)[7] *rice*
I will eat rice.

Wupwe ani ewe rayis
I:will eat(t) *the rice*
I will eat the rice.

Wupwe wun ewe kkonik
I:will drink(i) *the water*
I will drink some of the water.

Woleaian (Sohn 1975) shows similarities to this group. "Intransitive" verbs are found when there is no patient, "transitive" when there is a specific patient, and partitives are a special case. However, the treatment of partitives is different. In Marshallese and Trukese a partitive sense is expressed by an "intransitive" verb with a specific patient. In Woleaian, on the other hand, it is expressed by a "transitive" verb with a non-specific patient.

Woleaian

I be iul shal
I will drink(i) *water*
I will drink water.

I be iuliumi shal we
I will drink(t) *water the*
I will drink the water.

I be lag chuwaaiy filoowa
I will go buy(t) *bread*
I will go buy some bread.

Further, in Woleaian "transitivity" seems to have an association with perfectivity. Many of the intransitive verbs are formed by reduplication, which is a mark of imperfectivity in other Austronesian languages. Also, Sohn gives examples of question-word questions which with "transitive" verbs are translated as past tense and with "intransitive" verbs as present progressive. I have been unable to ascertain whether or not these facts are also true of Trukese and Marshallese.

Woleaian

Ye foori metta
he do(t) *what*
What did he do?

Ye ffoor[8] metta
he do(i) *what*
What is he doing?

Another group of Micronesian languages, which includes Kosrean (Lee 1975), Mokilese (Harrison 1976) and Ponopean (Rehg 1981), shows a slightly different pattern. In these languages verbs identified as intransitive can only appear with no patient or with a non-referential patient, which is said to be incorporated, since verbal aspectual suffixes are placed after it in Mokilese, and it shows phonological differences, having to do with vowel shortening (represented in the orthography by absence of the letter h after the vowel) in Ponopean. Whenever the patient is referential, a "transitive" verb must be used. It is not possible to have an "intransitive" verb with a referential patient. It appears, however, that in Ponopean at least it is possible to have a "transitive" verb with a non-referential patient.

Mokilese

Ngoah audoh-la rimeh-i
I *fill*(t):perf *bottle:this*
I filled this bottle.

Ngoah audohd rimeh-la
I fill(i) *bottle*:perf
I finished filling bottles.

Ponopean

I pahn perek-i lohs-o
I will unroll(t) *mat:that*
I will unroll that mat.

I pahn perek-los
I will unroll:mat
I will mat-unroll.

I pahn perek-i lohs
I will unroll(t) *mat*
I will unroll mats.

Ulithian (Sohn and Bender 1973) is described as having an "intransitive" verb form which is used with non-definite patients, and a "transitive" verb form which is used with definite patients. No further detail on the use of the verbs was provided, so it is not possible to determine whether it actually belongs in one of the two groups described above, or is a third subtype.

Relative clauses in some Micronesian languages show a constraint similar to the one found in many Western languages. As described above, in many Western languages the head of a relative clause must be the trigger of that clause. In Ponopean it appears to be impossible to form relative clauses on the agent when the patient is referential. Thus, the sentence "The dog that bit this child will be killed" cannot be directly translated as a single sentence. It is not clear what would happen if the patient were not referential. In Woleaian agent of a "transitive" verb is not listed as one of the possible functions within a relative clause for the head of that clause, while patient is. Again, there is no information about patient referentiality. The situation in Micronesian languages is not exactly parallel with Indonesian and Philippine languages (nor is it exactly clear), but there is certainly a similarity. In both the Micronesian and the Western languages, when there is a verbal affix indicating the semantic role of the patient, there is a restriction on relativisation. I have argued with reference to the Western languages that this restriction is a natural development in a transitivity marking system. Its presence among Micronesian languages gives further credence to this analysis.

It thus seems that many Micronesian languages show a strong correlation between transitive marking and patients that rank higher on the scale of individuated patients, being at least referential, and in some languages also specific or definite, and totally affected. There is also some evidence of a correlation between transitivity and perfectivity, in at least one language. Thus, transitivity in Micronesian languages is not determined by the number of arguments; it is discourse based. The most significant parameter seems to be patient status, and the indication of transitivity is the form of the verb, which reflects a historical process of suffixation. From the standpoint of a theory of discourse transitivity, it would be preferable to refer to the two Micronesian clause types as high transitivity and low transitivity, as I did in Philippine and Indonesian languages, rather than as transitive and intransitive, which is the more usual practice.

CENTRAL PACIFIC

Of the languages of the central Pacific region, Fijian shows the clearest correlation between transitivity and patient status. Unlike the Micronesian languages, Fijian shows clearly recognisable morphemes -(C)i, which is a reflex of *-i, -(C)a which is a reflex of *-i plus the third person singular object marker, and -(C)aki(ni) which is a reflex of *-aken. According to the standard analysis, in Fijian, as in Micronesian languages, these affixes co-occur with specific patients (Pawley 1973; Milner 1967).

Bauan Fijian

era saa tali magimagi
they aspect *plait*(i) *sennit*
They are plaiting sennit.

era saa tali-a na magimagi
they aspect *plait*(t) *the sennit*
They plait the sennit.

Recent discourse work (Schütz 1985) indicates that the facts are considerably more complex. In texts, the choice between suffixed and unsuffixed verbs does not depend on a polarised distinction between specific and non-specific. Rather, there is a "hierarchy of specificity" (p.387), and greater specificity correlates with use of the suffixed forms. This is similar to the situation in the Toba Batak and Tagalog texts, where appealing to an absolute notion of individuation yields an imperfect correlation with verb morphology. In all three languages, the structures available in the language are deployed by speakers, not by rule, but for effect.

According to Clark (1976) the Polynesian languages do not show a direct correlation between transitive marking and patient status. -(C)i and -(C)aki, reflexes of *-i and *-aken, are usually referred to as passive markers, and in some languages (e.g. Hawaiian, Tahitian) are used no more frequently than the passive in English. However, their use remains frequent in Maori (where it is termed "passive"), and in the Tongic and Samoic languages. In Tongic and Samoic, and in 19th century Maori texts, it is the preferred structure with agentive verbs and highly affected patients, and is infrequent with verbs that take an experiencer rather than an agent, where the patient is not particularly affected. Textual studies (Clark 1973) suggest a further correlation with perfective aspect and realis mood in Maori. Chung (1978) proposes that in Maori the use of the passive correlates with affected patients. She further points out that there is a restriction on relativisation in Maori, such that only the agent in an "active" clause or the patient in a "passive" clause can be relativised. In Samoan, *-i has traditionally been called a perfective marker. However, Chung (1978) notes that in texts it is most frequently used when the agent is an omitted generic, and in topicalisations, clefts and relative clauses on the agent. This last fact is surprising, as one would expect this form to be used in clefts and relative clauses on the patient, and suggests a need for further investigation of Samoan.

The situation in these Polynesian languages recalls some of the other parameters that Hopper and Thompson list. Certainly the use with agentive verbs, typically involving volition and kinesis, is a mark of high transitivity. So are perfectivity and realis, which may be significant in Maori. The Maori relativisation constraint is again reminiscent of the situation in some Micronesian and many Western Austronesian languages, where relativisation of the patient

requires patient trigger morphology within the relative clause. While these languages do not provide direct evidence of a correlation between transitive marking and patient status, they do show a correlation between transitive marking and discourse transitivity, as opposed to two-argument transitivity. This would be less likely if the Polynesian languages descended from a proto-language in which transitivity were defined in terms of number of arguments, and more likely if they were descended from a system such as that in Fijian, and the Micronesian languages, where transitivity is defined in terms of one of the other parameters.

Grace (1955, 1959) places Fijian and the Polynesian languages in the same subgroup, which he then combines with the nuclear Micronesian subgroup to form a higher order subgroup. Since patient status is so clearly associated with transitivity in both the Micronesian branch and the Fijian part of the Fiji-Polynesian branch, it seems safe to assume that this association was a part of the parent language, both to the entire group and to the Fiji-Polynesian branch. Was it present also in Proto-Polynesian? I argue that the evidence suggests that it was. If *-i and *-aken in PPN were not associated with high transitivity, the situation in the daughter languages would be inexplicable. However, if they were associated with high transitivity, the changes in the daughter languages (to affectedness of patient, perfectivity, etc.) are reasonable developments. Since, however, the daughter languages show such a variety of reinterpretations of the constraint, up to and including complete loss of it, it seems unlikely that the change took place in PPN. It is most likely that in PPN transitivity was defined as it is in Fijian and in the Micronesian languages, and that the changes took place after the breakup of PPN.[9]

MELANESIAN LANGUAGES

Many Melanesian languages show traces of *-i and *-aken, usually -(C)i and sometimes a form like -(C)aqi or -(C)a'i, or barring that, of the object agreement markers that co-occur with transitive suffixes in other languages. In Nguna (Schütz 1969), which is spoken in the New Hebrides, and in languages of the Solomons such as Florida (Ivens 1937), Ulawa (Clark 1973), Arosi (Capell 1971) and Kwara'ae (Deck 1934), "transitive" verbs (with either a transitive suffix, an object agreement marker or both) are used when the patient is present. Deck states that in Kwara'ae (and presumably in the other languages as well) it is presence in the discourse that is relevant, not presence in the clause itself.

Nguna

ku tatago-vi au noai naga pa munu
you ask:tr *I water that* imp *drink*
You asked me for water that you might drink.

In the descriptions of the languages mentioned above, the status of the patient does not appear to matter, only its existence. However, Simons (1979) points out that in Arosi and Bugotu not all clauses with patients have transitive suffixes. He argues on this ground that the suffixes are not transitive suffixes at all, but rather focus markers. However, he defines focus as clause-level topic, and neither defines topic, nor gives any evidence of why a certain NP should be topic rather than another. It is not clear from the limited data available to me on these two languages just what the conditions on the use of the suffix are, but a preliminary examination of one text for each language (Arosi - Capell 1971; Bugotu - Ivens 1933) suggests that patient individuation

may well be involved. In both texts most of the patients are referential, but there are two clauses in the Bugotu text, and one in the Arosi text where the patient is non-referential, as defined by DuBois (1980), and in these cases there is no transitive suffix. Further study of these languages might prove this to be significant.

Arosi

ma raru sio uri wou ...
and they pick type of fruit onwards
So they went along picking fruit ...

Bugotu

me suke bau
and (he was) mending nets

This may be true of other Melanesian languages as well. If so, it certainly seems a reasonable hypothesis that the association of transitive morphology and patient individuation, which is found both among Central Pacific-Micronesian languages, and (possibly) among Melanesian languages, should be attributed to the proto-language. And in fact, that is exactly what Pawley (1972, 1973) concluded, and his reconstruction was accepted by Clark (1973, 1976). They both reconstruct transitive marking with specific patients as far back as Proto-Eastern Oceanic, the ancestor of Nuclear Micronesian, Fijian, Polynesian, and most Melanesian languages from Florida and Guadalcanal islands eastwards. They do not suggest it for Proto-Oceanic, presumably for lack of evidence among the Western Melanesian languages. The lack of evidence reflects the lack of good descriptions of these languages. However, a comprehensive grammar of Manam, spoken in Papua New Guinea, has recently been published (Lichtenberk 1983), which offers interesting evidence.

In Manam, as in many Oceanic languages, there are transitive suffixes. When there is no patient, the "intransitive" form is used, and when there is a patient, the "transitive" is usually used. However, with certain classes of verbs the "intransitive" form is used "optionally with direct objects that are either non-specific or non-higher-animal, and they are used obligatorily if the direct object is both non-specific and non-higher-animal" (p.174).

Manam

boro zinzing u-rere-re
pig black I:like:redupl
I like black pigs.

boro zinzing nge u-rere-ta'-idi
pig black this I:like:tr:3pl
I like these black pigs. [adapted from p.175]

It appears that individuation of patient is significant for this language which, while Oceanic, is not Eastern Oceanic. This suggests that the association between transitivity and patient status may be reconstructable for POC. Without more data it will be difficult to confirm or disconfirm this hypothesis, but I believe that evidence from Western Austronesian languages throws light on this issue. Since patient status is important to transitivity marking in both Western and Oceanic languages, it seems quite likely that it can be reconstructed for the parent language of all of them.

CHAMORRO AND PALAUAN

Chamorro and Palauan, both spoken in geographical Micronesia, are excluded from Oceanic by all classifications. However, their exact relation with other Western Austronesian languages is unclear. Both show certain characteristics reminiscent of a typical trigger marking system, but are highly deviant in most regards.

Palauan (Josephs 1975) has a prefix me-ˬ-m-ˬ-o which is a reflex of *-um-, which is found with both transitive and intransitive actor trigger verbs. There is also a construction with a conjugated verb, using agent agreement prefixes, called the hypothetical form, which is found in conditional clauses and patient trigger constructions. In form this hypothetical verb is like patient trigger in a number of Western languages, in that it lacks the *-um- reflex, and makes use of agent clitics. According to Josephs, the Palauan patient trigger is relatively low in frequency, similar to the passive in English, and unlike a typical Western Austronesian language with a trigger marking system. It is, however, required when the patient in a relative clause is the head of that clause. This constraint, which is also found in Philippine and Indonesian languages with trigger marking systems, seems to connect the hypothetical verb form in a very clear manner with patient trigger. That is, not only is the hypothetical verb formally similar to patient trigger, but it shares some of the same syntactic constraints.

Palauan

A ngalEk a mE-nga Er a ngikEl
art *child* art act:*eat the* art *fish*
The child is eating the fish.

A nigkEl a lo-nga Er ngii a ngalEk
art *fish* art 3sg:*eat the by* art *child*
The fish is being eaten by the child.

A blai El lE-silsEbii a rEdil a blil a Toki
art *house the* 3sg:*burn* art *woman* art *house* art *Toki*
The house that the woman burned down was Toki's house.

Palauan has another construction (Wilson 1972), used for perfective verbs. Perfective marking may be used with "active" or hypothetical ("passive") verbs. Perfective verbs take object suffixes which cross-reference the person and number of the patient. However, this construction can only be used when the patient is specific. The perfective construction thus shows a similarity to Oceanic transitivity marking, even though there do not appear to be any reflexes of *-i or *-aken. I mentioned above that Oceanic languages often use object agreement markers as well as a transitive suffix with a transitive verb, and I have shown that transitive verbs in many Oceanic languages can be defined as verbs with individuated patients. The restriction of the perfective marker and object suffixes to situations where the patient is specific suggests that a correlation between specificity and object agreement may predate the separation of Oceanic languages, although it offers no evidence of a correlation with transitivity per se.

The situation in Chamorro is equally complex. Chamorro (Topping 1973) has a number of clause types, including a conjugated form which uses actor agreement prefixes, an actor trigger using *-um-, and a patient trigger using *-in-.[10]

Chamorro

Si Juan ha-li'e' i palao'an
art name 3s:*see* art *woman*
Juan saw the woman.

Si Juan l-um-i'ei i palao'an
art name af:*see* art *woman*
Juan saw the woman.

L-in-i'e' si Maria as Juan
pf:*see* art name art name
Juan saw Maria.

There is also a clause type for use with indefinite patients, which employs the prefix man-. This prefix, while not restricted to clauses with indefinite patients, is required with them. Topping (1973) refers to it as a detransitiviser, and Chung (1981) as an antipassive.

Chamorro

Guiya man-li'e' palao'an
he *see* *woman*
He saw a woman.

Chung (1981) points out that the choice among these forms appears to be conditioned at least in part by discourse transitivity. When a patient is indefinite, the form of the verb with man- is used. When the patient is definite, the choice between a conjugated verb and a patient trigger verb depends on the relative degree of individuation of the actor and patient, based on a hierarchy of pronoun > animate NP > inanimate NP. Chung found that conjugated forms could not be used if the patient was more highly individuated than the actor. Two alternative strategies were used to avoid this. Either patient trigger was chosen, or the man- prefix was used and the patient was marked as an oblique, not a direct argument.[11]

Chamorro

Ni-naketi i patgun ni manenghing
pf:*cause*:*cry* art *child* art *cold*
The cold made the child cry.

man-bisita i palao'an nu siha
visit art *woman* art *them*
The woman visited them.

Exceptions to this rule were cases where the actor NP was fronted either for topicalisation or question formation. Topicalisation and question formation are structurally and pragmatically similar to relativisation. In all three constructions, a referent is first established as topic, question word or head of relative clause, and then expanded on. Just as with relativisation on the actor in Philippine and Indonesian languages, in these cases also, patient trigger could not be used. Thus in Chamorro we find a correlation between individuation of the patient and trigger choice, parallel to the Oceanic correlation between individuation of the patient and transitive morphology. We also find a constraint similar to the constraint on relativisation previously mentioned.

SUMMARY OF EVIDENCE

A correlation between transitive verb morphology and presence or individuation of the patient is found in many modern Oceanic languages. A correlation between transitive morphology and individuation is reconstructed as far back as Proto-Eastern Oceanic, and can possibly be reconstructed for Proto-Oceanic.

Western trigger morphology has not generally been considered transitive morphology up till now. However, there is a strong correlation between patient trigger and patient individuation. Discourse studies of Batak and Tagalog show that patient trigger does not correlate with the prominence or thematicity of a patient in a text the way passive does in English. This suggests that the Western trigger marking system has the same discourse function as the Oceanic transitive marking, that of indicating high transitivity and patient salience, and that it is actually also transitivity morphology.

Thus, not only is transitive verb morphology found in language after language which, no matter what your subgrouping assumptions, represent most of the major branches of Malayo-Polynesian (or Extra-Formosan), but also it is a discourse definition of transitivity that recurs, with the same parameter, individuation of patient, as the deciding factor. What is the significance of this fact?

I can think of two possible explanations of this phenomenon. One is parallel development, and the other is shared inheritance. Parallel development is possible if either the change is so common among languages of the world as not to invite surprise if it recurs widely or if there is some reason in the structure of the proto-language that serves as a catalyst. Coincidental parallel development is an acceptable hypothesis only if the number of languages involved in the coincidence is fairly small.

The phenomenon is too widespread among Austronesian languages for me to find coincidental parallel development a likely choice. The marking of high transitivity (defined in terms of patient individuation) on the verb is not a sufficiently common phenomenon to be accepted as occurring without motivation. While it is true that Hopper and Thompson propose their definition of transitivity as universal, the language specific details of how it is expressed are not. A correlation between transitive marking and a number of parameters is predicted by the theory of transitivity. The choice of verb morphology as a means, and individuation of patient as the main parameter, are not. And I am at a loss to think of a motivating structure in the proto-language that could have induced such a drift. Furthermore, there are a great many languages that do not show the correlation. If there were truly a drift from some other state towards a correlation between transitivity marking and patient individuation, and some strong motivating factor in the parent language impelling this change, I would not expect the change to be so widely scattered. I would expect more languages to have drifted in that direction.

I find the shared inheritance model much more likely. This would require only one innovation of the correlation, at an extremely early stage. Those languages which no longer show the correlation are assumed to have widened their definition of transitivity to include all verbs with two arguments. This seems much more likely than that the original constraint was to mark transitivity only if two arguments were present, and that this constraint was narrowed in an almost identical manner by so many geographically separated languages. If narrowing had occurred, surely one would expect a greater variety of narrowing strategies than are attested.

Language contact, while often a possible candidate for explaining areal phenomena, seems to be unlikely in this case. Although the phenomenon is found throughout the region, there are considerable gaps. The most notable one appears to be in the area of Melanesia, but there are many other smaller ones as well. If the correlation between transitivity marking and patient individuation spread by contact, it seems strange that it is found predominantly but sporadically in the west and the east, and not in the central area.

THE MORPHOLOGY OF TRANSITIVITY

I have attempted to reconstruct a function for verbal morphology which could produce the current variety of functions among the Extra-Formosan languages. I have not commented on the actual morphology used. I believe that my functional reconstruction is compatible with most of the morphological reconstructions that have been proposed, with the possible exception of the Philippine hypothesis.

The distribution of the functions among the modern languages is as follows (for evidence of the use of trigger marking morphology for nominalisation, see Starosta, Pawley and Reid 1982; for an explanation of this double function, see below):

Philippine languages use reflexes of *-um-, and either *-in-, *-en, *-an or *Si- to distinguish low and high transitivity respectively. These morphemes are also used as nominalisers. *-um- forms agentive nouns, *-in- and *-en form affected nouns, *-an forms locative nouns and *Si- forms instrumental nouns. Reflexes of *-i are found as frozen nominal and pronominal prefixes, and occasionally as a locative preposition. They mark dependent forms and imperatives in a number of languages. However, it does not appear to be a productive part of the trigger marking system. *-aken is not directly reflected in the Philippines, but -a is found in imperatives and dependent forms.

Indonesian languages use reflexes of *-um-, *-in-, *-an and *Si- as nominalisers. The reflexes of *-um- and *-in- are also used to distinguish low and high transitivity respectively, and in some languages the reflexes of *-an and *Si- may also be used. They also use reflexes of *-i and *-aken in a transitivising function, but not in a way related to distinguishing high and low transitivity. These morphemes increase the valence of a verb, making one-argument verbs into two-argument verbs, and two-argument verbs into three-argument verbs.[12]

In Oceanic languages, reflexes of *-i and *-aken mark high transitivity, while reflexes of *-um-, *-in-, *-an and *Si- are nominalisers. Table 5 represents the distribution of functions described above.

Table 5

	OCEANIC	INDONESIAN	PHILIPPINE
-i/-aken	scalar transitivity	valence increasing	not productive
-um-/-in-/ *Si-/*-an	nominalisers	scalar transitivity (nominalisers)	scalar transitivity (nominalisers)

The current situation could have developed from a number of different earlier configurations, as suggested by the competing hypotheses about the development of the trigger marking system. One can hypothesise how this might have happened, although perhaps not why it did.

The process could begin with a system in which some morphemes in PMP had two functions, high transitivity and nominalising, as is the case now in Indonesian and Philippine languages. This is consistent with the Toba Batak hypothesis. In POC these morphemes became restricted to nominalising, while the valence-increasing marker moved in to take over the high transitivity function. To get to the Philippine type, the only necessary change is loss of productivity of the valence-increasing morpheme.[13]

Table 6

PMP/INDONESIAN		OCEANIC	PHILIPPINE
Scalar Trans/Nom	>	Nom	Scalar Trans/Nom
Valence	>	Scalar Trans	non-productive

It is also possible to imagine a system where original PMP nominalisers became scalar transitivity markers while the original scalar transitivity markers either became valence-increasing markers or were lost. This scenario goes from a system like the Oceanic one to systems like the Indonesian and Philippine types respectively, as both the Oceanic hypothesis and the Formosan hypothesis suggest.

Table 7

PMP/OCEANIC		INDONESIAN	PHILIPPINE
Nom	>	Scalar Trans/Nom	Scalar Trans/Nom
Hi Trans	>	Valence	non-productive

It is more difficult to imagine that a system with morphemes that were simultaneously scalar transitivity markers and nominalisers, but without any other productive set of affixes, like the Philippine system, could change into both the Indonesian and the Oceanic types. This is especially the case since *-i and *-aken are reconstructed for PAN, and *-i is attested in Formosan languages. It would be absurd to suggest that they did not exist in some productive function in Proto-Extra-Formosan. They could not reasonably have fossilised and then been resurrected independently in Oceanic and Indonesian languages.

The combination of scalar transitivity and nominalisation in the same morpheme at first seems puzzling from a discourse standpoint, since nominalisations are typically low transitivity. This presents a problem for the hypotheses suggested above. To be acceptable, any hypothesis must be accompanied by a reasonable explanation for this combination. Starosta, Pawley and Reid give a formal explanation, suggesting that in PAN the original nominalisers *-um-, *-in-, *-en, *-an and *Si-, were reanalysed as part of the trigger marking system due to formal similarities with clauses using *-i and *-aken.

The theory of transitivity requires a discourse-based explanation of why this reanalysis could occur. I would like to suggest the following: while it is true that nominalisations per se are typically low in transitivity, the content of an individual nominalised phrase may be either low or high in transitivity. It is also not true that the entire trigger marking system marks high transitivity.

In particular, actor trigger (*-um-) marks low transitivity, while patient trigger (*-in-) marks high transitivity. It is believed that *-in- functioned as a marker of perfectivity in PAN. Perfectivity is characteristically associated with high transitivity. Thus, the reanalysis is not as bizarre as it first appears. The differing degrees of transitivity within individual nominalised phrases came to mark differing degrees of transitivity of clauses.

I thus propose this reconstruction as an addition to, rather than as a replacement of current reconstructions of PMP verbal morphology. Other reconstructions have focused on the forms of the morphemes, and their syntactic distribution. My interest is in discovering the discourse function behind this syntactic distribution.

TRANSITIVITY: PMP OR PAN?

The last question I wish to address is the relevance of this reconstruction to PAN itself. In other words, is the correlation between transitivity and patient individuation an innovation at the Proto-Extra-Formosan level (or perhaps the Proto-Amis-Extra-Formosan level), or is it a characteristic of PAN? I am not sufficiently familiar with Formosan languages to answer this question in a definitive way. Starosta, Pawley and Reid (1982), state that in PAN definite patients were always subjects. In my terminology, this would mean that whenever a definite patient occurred, transitive (or trigger marking, or focus) morphology would be used. If this is indeed the case, then the correlation I have proposed can appropriately be pushed back to the PAN level. To be certain, one would wish to have descriptions of Formosan languages which exhibited this same correlation.

NOTES

1. I would like to thank Preston Ashbourne, Niko Besnier, Susanna Cumming, Pamela Munro, Ellen Rafferty, Paul Schachter and Sandra Thompson for their helpful comments, and Wilson Manik for providing the Toba Batak data.

2. Subject and object are used here as naive, pretheoretical terms.

3. There is some uncertainty about the exact shape of the second suffix. I am following Starosta, Pawley and Reid's (1982) reconstruction.

4. The labels AT and PT refer to actor trigger and patient trigger respectively.

5. The use of ay marks a fronted argument.

6. The particle do probably indicates emphasis of some sort.

7. (i) and (t) designate respectively the "intransitive" and "transitive" forms of verbs.

8. Note the partial reduplication on the "intransitive" verb.

9. The implications of this proposal for the controversy over the direction of drift in Polynesian languages, accusative to ergative or ergative to accusative, should prove interesting.

10. Definite nouns in Chamorro are preceded by a variety of articles, depending on syntactic function and semantic class.

11. Chung did not discuss the use of actor trigger.

12. This is a simplification of the actual situation, but is sufficiently detailed for the purposes of this paper.

13. There is a problem here, in that the traces of these two morphemes in Philippine languages do not reflect a valence increasing function.

REFERENCES

ANCEAUX, J.C.

1952 *The Wolio language*. The Hague: Martinus Nijhoff.

BENDER, Byron W.

1969 *Spoken Marshallese*. PALI Language Texts: Micronesia. Honolulu: University of Hawaii Press.

BLOOMFIELD, Leonard

1917 *Tagalog texts with grammatical analysis*. Urbana, Illinois: University of Illinois Press.

CAPELL, A.

1971 *Arosi grammar*. *PL*, B-20.

CHUNG, Sandra

1978 *Case marking and grammatical relations in Polynesian*. Austin: University of Texas Press.

1981 Transitivity and surface filters in Chamorro. In J. Hollyman and A. Pawley, eds *Studies in Pacific languages & cultures in honour of Bruce Biggs*, 311-332. Auckland: Linguistic Society of New Zealand.

CLARK, Ross

1973 Transitivity and case in Eastern Oceanic languages. *Oceanic Linguistics* 12:559-605.

1976 *Aspects of Proto-Polynesian syntax*. *Te Reo* Monographs. Auckland: Linguistic Society of New Zealand.

CONSTANTINO, Ernesto

1971 *Ilokano reference grammar*. Honolulu: University of Hawaii Press.

COOREMAN, A., B. FOX and T. GIVÓN

1984 The discourse definition of ergativity. *Studies in Language* 8: 1-34.

DECK, Norman C.

1934 *A grammar of the languages spoken by the Kwara'ae people of Mala, British Solomon Islands*. New Plymouth, New Zealand: Thomas Avery.

DUBOIS, John

1980 Beyond definiteness: the trace of identity in discourse. In W.L. Chafe, ed. *The pear stories: cognitive, cultural, and linguistic aspects of narrative production*, 203-274. Norwood, NJ: Ablex.

FLORA, JoAnn

1974 The Palauan */m/* affix. *Oceanic Linguistics* 13:213-227.

FOLEY, William A.

1976 Comparative syntax in Austronesian. Ph.D. dissertation, University of California, Berkeley.

GIVÓN, Talmy

1979 *On understanding grammar*. New York: Academic Press.

GRACE, George W.

1955 Subgrouping of Malay-Polynesian: a report of tentative findings. *American Anthropologist* 57:337-339.

1959 *The position of the Polynesian languages within the Austronesian language family*. Baltimore: Waverly Press.

HALIM, Amran, Lois CARRINGTON and S.A. WURM, eds

1982 *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*. *PL*, C-75.

HARRISON, Sheldon P.

1976 *Mokilese reference grammar*. PALI Language Texts: Micronesia. Honolulu: University of Hawaii Press.

HARVEY, Mark

1982 Subgroups in Austronesian. In Halim, Carrington and Wurm, eds 1982: 47-99.

HOHEPA, Patrick W.

1969 The accusative-to-ergative drift in Polynesian languages. *Journal of the Polynesian Society* 78/3:295-329.

HOPPER, Paul J.

1979a Some discourse sources of ergativity. *Working Papers in Linguistics, University of Hawaii* 11/2:137-153.

1979b Aspect and foregrounding in discourse. In T. Givón, ed. *Syntax and semantics*, vol.12: *discourse and syntax*, 213-241. New York: Academic Press.

HOPPER, Paul J. and Sandra A. THOMPSON

1980 Transitivity in grammar and discourse. *Language* 56:251-299.

HORNE, Elinor Clark

1961 *Beginning Javanese*. New Haven: Yale University Press.

IVENS, W.G.

1933 A grammar of the language of Bugotu, Ysabel Island, Solomon Islands. *BSOAS* 7:141-177. London: Society of Asiatic Studies.

1937 A grammar of the language of Florida, British Solomon Islands. *BSOAS* 8:1075-1110. London: Society of Asiatic Studies.

JOSEPHS, Lewis S.

1975 *Palauan reference grammar.* PALI Language Texts: Micronesia. Honolulu: University of Hawaii Press.

LEE, Kee-dong

1975 *Kusaiean reference grammar.* PALI Language Texts: Micronesia. Honolulu: University of Hawaii Press.

LICHTENBERK, Frantisek

1983 *A grammar of Manam. Oceanic Linguistics* Special Publication No.18. Honolulu: University of Hawaii Press.

MILNER, G.B.

1967 *Fijian grammar.* Fiji: Government Printing Department.

NAYLOR, Paz B., ed.

1979 *Austronesian studies: papers from the Second Eastern Conference on Austronesian Languages.* Michigan Papers on South and Southeast Asia No.15. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.

PAWLEY, Andrew

1972 On the internal relationships of the Eastern Oceanic languages. In R.C. Green and M. Kelly, eds *Studies in Oceanic culture history* 3, 1-142. Pacific Anthropological Records 13. Honolulu: Department of Anthropology, Bernice Pauahi Bishop Museum.

1973 Some problems in Proto-Oceanic grammar. *Oceanic Linguistics* 12:103-188.

PAWLEY, Andrew and Lawrence A. REID

1979 The evolution of transitive constructions in Austronesian. In Naylor, ed. 1979:103-130.

POEDJOSOEDARMO, Gloria

1976 Hypotese perkembangan sintaksis bahasa Jawa. *Bahasa dan Sastra* 11/4: 2-22. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

REID, Lawrence A.

1982 The demise of Proto-Philippines. In Halim, Carrington and Wurm, eds 1982:201-216.

REHG, Kenneth L.

1981 *Ponopean reference grammar.* PALI Language Texts: Micronesia. Honolulu: University of Hawaii Press.

SCHACHTER, Paul

1976 The subject in Philippine languages: topic, actor, actor-topic, or none of the above? In Charles Li, ed. *Subject and topic*, 493-518. New York: Academic Press.

SCHACHTER, Paul and Fe OTANES

1972 *Tagalog reference grammar.* Berkeley, CA: University of California Press.

SCHÜTZ, Albert J.

1969 *Nguna grammar.* Honolulu: University of Hawaii Press.

1985 *The Fijian language.* Honolulu: University of Hawaii Press.

SIMONS, Gary

1979 The verbal sentence in Arosi: a reinterpretation of Eastern Oceanic Sentence structure. In Naylor, ed. 1979:131-151.

SOHN, Ho-min

1975 *Woleaian reference grammar.* PALI Language Texts: Micronesia. Honolulu: University of Hawaii Press.

SOHN, Ho-min and Byron W. BENDER

1973 *A Ulithian grammar. PL,* C-27.

STAROSTA, Stanley, Andrew K. PAWLEY and Lawrence A. REID

1982 The evolution of focus in Austronesian. In Halim, Carrington and Wurm, eds 1982:145-170.

SUGITA, Hiroshi

1973 Semitransitive verbs and object incorporation in Micronesian languages. *Oceanic Linguistics* 12:393-406.

THOMPSON, Sandra A.

forthcoming The passive in English: a discourse perspective. To appear in Robert Channon and Linda Shockey, eds *In Honor of Ilse Lehiste: Ilse Lehiste Suhendusteos.*

TOPPING, Donald M.

1973 *Chamorro reference grammar.* PALI Language Texts: Micronesia. Honolulu: University of Hawaii Press.

TUUK, H.N. van der

1971 *A grammar of Toba Batak.* The Hague: Martinus Nijhoff. (Translation of *Tobasche Spraakkunst,* 1864.)

VERHAAR, John W.M.

forthcoming Syntactic ergativity in contemporary Indonesian. To appear in McGinn, ed. *Proceedings of the Third Eastern Conference on Austronesian Languages.*

WILSON, Helen

1972 *The phonology and syntax of Palauan verb affixes. Working Papers in Linguistics, University of Hawaii* 4/5.

WOLFF, John U.

1979 Verbal morphology and verbal sentences in Proto-Austronesian. In Naylor, ed. 1979:153-167.

1980 *Beginning Indonesian.* Ithaca: Cornell University Press.

WOUK, Fay

1984 Scalar transitivity and trigger choice in Toba Batak. In P. Schachter, ed. *Studies in the structure of Toba Batak*, 195-219. UCLA Occasional Papers in Linguistics. Los Angeles: University of California.

forthcoming a Trigger choice in Batak and Tagalog. To appear in *Studies in Language*.

forthcoming b Verbal morphology and the discourse structure of Indonesian. To appear in *Proceedings from the Fourth Eastern Conference on Austronesian Languages*.

Wouk, F. "Transitivity in Proto-Malayo-Polynesian and Proto-Austronesian". In Geraghty, P., Carrington, L. and Wurm, S.A. editors, *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics.* **C-93:133-158**. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.133*

THE PRESENCE AND ABSENCE OF *meN-*: THE INDONESIAN TRANSITIVE VERBS

Bambang Kaswanti Purwo

The Indonesian transitive verbs can be classified into two major forms: the meN- form and the non-meN- form.[1] Consider the diagram below.

	(A) meN-	(B) non-meN-	
		(B1) zero	(B2) di-
1st person	saya/aku mengambil	saya/aku ambil kuambil	
2nd person	engkau/kamu mengambil	engkau/kamu ambil kauambil	
3rd person	dia/ia mengambil	dia ambil	diambilnya diambil olehnya
nominal	si Dul mengambil		diambil (oleh) si Dul

[:ambil *to take*; aku/-ku/saya *I*; engkau/kau-/kamu *you*; dia/ia/-nya *he, she*; oleh *by*]

The verbal form (A) has traditionally been called an active verb, (B) a passive verb. Scholars like Alisjahbana (1954), Slametmuljana (1969), and Ramlan (1977) use the terms "active" and "passive" in their linguistic description, whereas scholars like Mees (1950), Fokker (1951), Kähler (1956), Teeuw (1971), and Danusugondo (1976) prefer not to use the terms "active/passive" at all (instead, some have suggested that (A) be called a subjective construction, and (B) an objective construction). Samsuri (1976) claims that use of the name "passive" is grounded on a Graeco-Latin analysis.

Recent studies (Cartier 1979, Rafferty 1982) have introduced the term ergative along with the term passive, for the (B) verbal form. Verhaar (1983) further proposed that (A) can be active as well as antipassive, while (B) can be passive as well as ergative, depending on certain contextual (discourse) constraints.

Chung (1976) was the first to propose that (B1) be distinguished from (B2); the former she called "object preposing", the latter "canonical passive". Although I do not agree with her arguments (Kaswanti 1984), the distinction

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 159-170. *Pacific Linguistics*, C-93, 1986.

between the two forms is important. In older narrative Malay one does not find the verbal form (B1), and in the conversations in 16th century Malay compiled by Frederick de Houtman (Lombard, ed. 1970), instead of (B2) one can only find (B1). There seems to be a complementary distribution between the use of (B1) for the spoken language and the use of (B2), especially di-/-nya, for the written language in older Malay. While the two forms are found in the written language of present-day Indonesian, each shows its own characteristics (Kaswanti 1983a, 1984).

The present paper is an attempt to describe the difference between the meN- form and the non-meN- form in terms of the verbal 'nuance'. By 'verbal nuance' I mean the way the verb most frequently expresses the relationship of the speech participants, either to one another or to the verb. In English there are certain verbs which, in certain types of constructions, most often have first person subjects along with the (simple) present tense, as exemplified by the following:

(1)a. *I ASK you, therefore, not to be discouraged because of my sufferings for you* [....] (Ephesians 3:13)

b. *And I PRAY that you, being rooted and established in love, may have power* [....] (Ephesians 3:17)

c. [...] *I URGE you to live a life worthy of the calling you have received.* (Ephesians 4:1)

These types of sentences have been labelled "performative" after Austin, as quoted by Ross (1970:222). They correspond to (or at least are the closest analogy to) the verbal nuance I am attempting to describe below. In Indonesian, whether the nuance is "performative" or "non-performative" is detectible in the absence or the presence of the meN- form. The term "performative", although it is the closest one available, is not in fact appropriate for the present purpose. To provide a better picture of what I mean by verbal nuance here I am using the following parameters as well: telic/atelic, durative/punctual. The types of verbs (e.g. verbs of saying, verbs of consciousness, verbs of perception) are also taken into consideration.

The parameters 'telic' and 'atelic' can be useful to articulate the distinction between the meN- form and the non-meN- form, but they are not applicable to verbs of saying (e.g. mengatakan *to say*; menanyakan *to ask*; menceritakan *to tell*). By contrast, the terms 'durative' and 'punctual', despite the usefulness for the description of the verbal nuance, fail to apply to verbs of consciousness (e.g. mengira *to think*; menyangka *to consider*) and to verbs of perception (e.g. melihat *to hear*) which are non-volitional. The parameters 'durative' and 'punctual' are only applicable to volitional verbs; verbs of consciousness are non-volitional, whereas verbs of perception can be volitional as well as non-volitional.

If one endeavours to see the distinction between the meN- form and the non-meN- form from the English translation, one might come across the difference between the past tense and the non-past tense, respectively. Consider (2), where the meN- form coincides with the past tense, the non-meN- form with the non-past.[2]

(2)a. Saya MENDOAKAN supaya kamu menang, tetapi ternyata kamu kalah.
*doakan
I PRAYED that you would win, but you lost.

b. Saya *mendoakan supaya kamu menang nanti.
DOAKAN
I PRAY that you will win.

The endeavour to adopt the tense distinction, such as in English, however, is proven useless, when one examines verbs of consciousness. Both English present and English past tense are translatable either into the meN- form (3) or into the non-meN- form (4).

(3) *Why do/did you think so?*
Mengapa kamu MENGIRA begitu?
why you think so

(4) *Do you think he is not serious?/Did you think he was not serious?*
Apakah kauKIRA dia tidak bersungguh-sungguh?
QWQ *you-think he not serious*

The following two pairs of sentences are in the English past, but (5) is expressed in Indonesian using the meN- form, while (6) uses the non-meN- form.

(5) *Yes, I thought at first that he was a communist.*
Ya, semula saya MENGIRA dia komunis.
yes before I think he communist

(6) *I thought he was a communist, but he is not.*
Saya KIRA dia komunis, ternyata bukan.
I think he communist not

When we examine the difference between the meN- form and the non-meN- form in examples (2) through (6), what is in common in the data is that the meN- form is telic, while the non-meN- form is atelic. The same parameters can also be found in the verbs of perception. Consider (7) and (8).

(7) *Have you ever seen/Did you see that movie?*
Pernahka kamu MELIHAT film itu?
ever you see film that

(8) *Do you see that?*
KauLIHAT itu?
you-see that

It is, however, not the telicity which matters in a context like (9) and (10). The verbs of perception in both (9) and (10) are telic.

(9) "Aku tak pernah MELIHAT ayahmu. Di mana dia selama ini?" tanya
I never see father-your in where he so far ask

Tody tiba-tiba. "Sering bepergian", jawab, Irawati. (Ashadi Siregar, 69)
suddenly often travel answer
"I never see your father. Where is he these days?" Tody asked suddenly.
"He often travels", Irawati answered.

(10)a. Sudah lama si Dul tidak kelihatan di gereja.
already long not be seen at church
It's been a long time since we saw Dul at church.

b. Tapi saya LIHAT dia kemarin ke gereja.
but I see he yesterday to church
But I saw him go to church yesterday.

The telicity is not proper either with verbs of saying. The terms 'durative' and 'punctual' are more appropriate, instead, to describe the distinction between the meN- form and the non-meN- form in (11) and (12): the meN- form is durative, and the non-meN- form is punctual.

(11) *Why do you always tell him the same story over and over?*
Mengapa senantiasa kamu MENCERITAKAN kisah yang itu-itu juga kepadanya?
why always you tell story LIG. *that-that* EMPH *to him*

(12) *Why did you tell this to me?*
Mengapa kamu CERITAKAN hal ini kepadaku?
why you-tell thing this to-me

The terms 'durative' and 'punctual', however, are not applicable to describe the distinction between the meN- form and the non-meN- form in examples (2) through (10). We need to find other parameters to explicate the difference between the two verbal nuances. As examples (3) through (6) illustrate, the English translation is proven not helpful. The following pair of sentences more obviously indicates that the meN- form and the non-meN- form are hard to discriminate between through the English gloss alone.

(13) Saya hanya akan MENGATAKAN dua hal ini: [....]
I only will say two thing this
I'm only going to say these two things: "[....]"

(14) Akan kuKATAKAN kepada si Dul bahwa kamu sakit.
will I-say to that you ill
I'm going to tell Dul that you are ill.

In order to understand the difference between (13) and (14) one has to consider the speech act of the two utterances. (14) is a 'propositive'[3] statement, while (13) is not. (14) is uttered in a context where the speaker is proposing something to the addressee; the English translation may also look like this: *Let me tell Dul that you are ill.* The use of the non-meN- form implies that the act of saying is performed for the benefit of the addressee. By contrast, the use of the meN- form in (13) clearly shows that the act of saying is for the benefit of the speaker rather than the addressee; the English translation may also read as follows: *I'm coming here to tell you the following two things.* In other words, (14) is more concerned with or more directly affects the addressee, when compared to (13).

Imperatives are typical constructions which express an act of imposing something directly on the addressee. In such constructions the meN- prefix of the transitive verbs is always dropped. The meN- form is retained only when the verb is intransitive.[4] Compare (15) to (16).

(15) *Membacalah buku itu!
BACAlah
Read that book!

(16) MENEPILAH!
Step aside!

The intransitive verb like the one in (16) has only one possible verb form, unlike the transitive verbs, which have the possibility of either dropping or retaining the meN- form. When the imperative construction is in the negative,

both meN- and non-meN- forms can be used. The use of the meN- form in negative imperatives indicates a lesser degree of command; the command is less direct and sounds more polite with meN- (Kaswanti 1983b).

(17) Jangan MEMBACA buku itu!
BACA
Don't read that book!

The directness and indirectness of the speech act is also remarkable in the exhortative constructions. The utterance has a more immediate impact when meN- is absent from the verb; with meN- the utterance would sound more polite and indirect. Compare (18) with meN- and (19) without it.

(18) Bagaimana kalau sekarang sebaiknya kita MELIHAT apartemen itu sebentar?
how if now better we see apartment that a moment
What about having a look at that apartment first?

(19) Mari kita LIHAT sebentar (apartemen itu).
let we see a moment apartment that
Let's have a look at that apartment.

Returning now to examples (2), (2b) indicates that the speaker is 'performing' the utterance, whereas in (1a) the speaker is 'narrating'; (2b) is 'performative', (2a) is 'narrative'. The opposition whether the verb is acted out (put on stage) or whether the verb is reported is also remarkable in (6) and (5), respectively. When we examine (3), (4), (9), and (10), we can see that the use of non-meN- in (4) and (10) demonstrates that the speaker is arguing for something; the non-meN- verb is foregrounded. By contrast, with the meN- form the speaker is commenting on the previous discourse (in the case of (3)), and is setting out the scene (in the case of (9)); the meN- verb is backgrounded. 'Performative' statements are marked by their vividness, directness, and foregroundedness, while 'narrative' statements are marked by their indirectness and backgroundedness.

The parameters 'performative' and 'narrative' are also found in constructions other than meN- and non-meN-. When the matrix verb is of the performative type (like menganjurkan *to recommend*; mengusulkan *to propose*), the subordinative clause is introduced with supaya *that* (20). On the other hand, when the matrix verb is of the narrative type (e.g. mengatakan *to say*; menceritakan *to tell*), the subordinative clause is introduced by bahwa *that* (21).

(20) Dia menganjurkan SUPAYA saya segera pergi ke dokter.
He recommended THAT I go to see the doctor now.

(21) Dia mengatakan BAHWA dia tidak pernah pergi ke dokter.
He said THAT he never went to see a doctor.

Another example which shows that a performative context may be distinguished from a narrative one is found in the case of the possible and the impossible deletion of the second person pronoun. The second person pronoun can be deleted in a performative context (23), but it cannot be omitted in a narrative context (22).

(22) Kapan *(kamu) pergi ke sana?
when you go to there
When did you go there?

(23) (Kamu) mau pergi ke sana sekarang?
you will go to there now
Are you going there now?

The constraint for the choice of meN- instead of non-meN- as has been described above is pragmatically conditioned. There seems to be some syntactic pressure for such a choice. The co-occurrence of modals with transitive verbs may cause the verbs to take the meN- form. To illustrate this point let us take an extreme example. There are certain verbs (very restricted in number) which, although they take an object (which makes them transitive-like), in a construction like (24) they are very unlikely to take the meN- form. (To mention some, they are ingat *to remember*; kenal *to know, recognise*; tahu *to know, understand*; lupa *to forget*; tanya *to ask (a question)*; usul *to propose (an idea)*.) These verbs, when preceded by modals, require the presence of the meN- form. Compare (24) to (25).

(24) Saya *mengingat wajah itu.
INGAT
I recognise that face.

(25) Saya tidak dapat MENGINGAT wajah itu lagi.
I cannot recognise that face any more.

Pragmatic consideration, however, overrides syntactic constraints. It is not always the case that the non-meN- form is ruled out whenever the modal is present. Consider (26) and (27).

(26) Apakah hárus // dia MENGAMBIL buku itu?[5]
QWQ *must he take book that*
Is it a "must" that he took that book?

(27) Apakah harus dia AḾBIL // buku itu?
Does he have to take that book?

The modal gets an extra emphasis in (26), whereas in (27) the emphasis falls upon the main verb. The absence of meN- indicates that the verb is foregrounded; the verb is contrasted (i.e. the act of ambil *to take* is in contrast, for example, with the act of beli *to buy* (e.g. as can be implied in the context of (26)): *Should he just take the book or should he buy it?*).

There is one remaining problem with respect to the pragmatic constraint for a construction like (27). Such a construction is only permissible when the patient, which is an afterthought topic, is third person. When the afterthought topic is second person, the construction containing the verb without the meN- form is not acceptable (29).

(28) Apakah haru dia MENGANTAR {orang itu? / *man that* / kamu? / *you*}
QWQ *must he go with*
Is it a must that he accompanies that man/you?

(29) Apakah harus dia ANTAR // {orang itu? / *man that* / *kamu? / *you*}
QWQ *must he accompany*
Does he have to accompany that man/you?

NOTES

1. The work for this paper was completed during my one year stay (1983-84) at The Institute for Advanced Study, Princeton.

2. See the appendix for more data on the recast contrast between the meN- and the non-meN- form.

3. Javanese grammarians (such as Berg 1937:2, 112; Bezemer 1931:49; Jansz 1893: 348ff; Prijohoetomo 1937:100ff; Roorda 1855:319) call -(n)é in a Javanese sentence like (i) and (ii) 'propositive'; such a construction is paraphrasable in English into *Let me DO that.*

   (i) Tak cobané.
   1st pers. *try*-PROP.
   Let me try that.

   (ii) Lawang kuwi tak tutupé.
   door that 1st pers. *close*-PROP.
   Let me close that door.

   Propositive sentences are only for first person, for positive statements, and for non-past tense.

4. With transitive verbs, meN- is used when the object is generic or unspecified.

   (iii) Mencurilah sebanyak-sebanyaknya.
   steal-PART. *as much as possible*
   Steal [unspecified object] *as much as you can.*

   One of the Ten Commandments, "Don't steal!" (or "Thou shalt not steal") is translatable into Indonesian with the meN- form: Jangan mencuri!

5. The modal mau in (iv) can either mean *will* [future] or *be willing to* [intention], but (v) can only have the second meaning while (vi) has the first.

   (iv) Dia mau mengambil buku itu sekarang.
   he take book that now
   He is going to/willing to take the book now.

   (v) Apakah mau // dia MENGAMBIL buku itu?
   QWQ *he take book that*
   Is he willing to take that book?

   (vi) Apakah mau dia AMBIL // buku itu?
   QWQ *he take book that*
   Is he going to take that book?

REFERENCES

ALISJAHBANA, St. Takdir

1954 *Tatabahasa baru Indonesia.* Djakarta: Pustaka Rakjat.

BERG, C.C.

1937 *Bijdrage tot de kennis der javaansche werkwoordsvormen. Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië* 95.

BEZEMER, T.J.

1931 *Beknopte javaansche grammatica.* Zwolle: W.E.J. Tjeenk Willink.

CARTIER, Alice

1979 De-voiced transitive verb sentences in formal Indonesian. In Frans Plank, ed. *Ergativity: towards a theory of grammatical relations*, 161-183. New York: Academic Press.

CHUNG, Sandra

1976 On the subject of two passives in Indonesian. In Li, ed. 1976:57-98.

COMRIE, Bernard

1976 *Aspect: an introduction to the study of verbal aspect and related problems.* Cambridge: Cambridge University Press.

DANUSUGONDO, Purwanto

1976 *Bahasa Indonesia. Indonesian for beginners* 2. Sydney: Sydney University Press.

FOKKER, A.A.

1951 *Inleiding tot de studie van de Indonesische syntaxis.* Djakarta: J.B. Wolters.

GAREY, Howard B.

1957 Verbal aspect in French. *Language* 33:91-110.

GONDA, J.

1949 Over Indonesische werkwoordsvormen I. *Bijdragen tot de Taal-, Land- en Volkenkunde* 105:333-379.

JACOBS, Roderick A. and Peter S. ROSENBAUM, eds

1970 *Readings in English transformational grammar.* Waltham, Mass.: Ginn.

JANSZ, P.

1983 *Javaansche spraakkunst: geheel omgewerkte, vermeerde en verbeterde uitgave van de kleine javaansche spraakkunst.* Semarang: G.C.T. van Dorp.

JESPERSEN, Otto

1951 *The philosophy of grammar.* London: George Allen & Unwin.

KÄHLER, Hans

1956 *Grammatik des Bahasa Indonesia.* Wiesbaden: Harrossowitz.

KASWANTI PURWO, Bambang

1983a The verbal aspect: a case of the Indonesian passives. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.4:239-250. *PL*, C-77.

1983b Imperative constructions: a case in Indonesian. MS, Institute for Advanced Study, Princeton.

1984 *Deiksis dalam bahasa Indonesia.* Seri ILDEP. Jakarta: Balai Pustaka.

KUNO, Susumu

1976 Subject, theme, and the speaker's empathy - a reexamination of relativization phenomena. In Li, ed. 1976:417-444.

LI, Charles N., ed.

1976 *Subject and topic*. New York: Academic Press.

LOMBARD, Denys, ed.

1970 *Le "Spraeck ende woord-boek" de Frederick de Houtman: première méthode de malais parlé (fin du XVIe s.)*. Paris: Ecole Français d'Extrême-Orient.

McCUNE, Keith

1979 Passive function and the Indonesian passive. *Oceanic Linguistics* 18:2:119-169.

MEES, C.A.

1950 *Tatabahasa Indonesia*. Bandung: G. Kolf.

MOURELATOS, Alexander P.D.

1978 Events, process, and states. *Linguistics and Philosophy* 2:415-434.

PRIJOHOETOMO, M.

1937 *Javaansche spraakkunst*. Leiden: E.J. Brill.

RAFFERTY, Ellen

1982 *Discourse structures of the Chinese Indonesian of Malang*. NUSA 12. Jakarta: Badan Penyelenggara Seri NUSA.

RAMLAN, M.

1977 *Masalah aktif-pasif dalam bahasa Indonesia*. Yogyakarta: Lembaga Penelitian Universitas Gadjah Mada.

ROSS, John Robert

1970 On declarative sentences. In Jacobs and Rosenbaum, eds 1970:222-227.

ROORDA, T.

1855 *Javaansche grammatica, benevens een leesboek tot oefening in de javaansche taal*. Amsterdam: Johannes Muller.

SAMSURI

1976 Verbal prefix *di-* in bahasa Indonesia. MS, IKIP Malang.

SLAMETMULJANA

1969 *Kaidah bahasa Indonesia*. Ende: Nusa Indah.

SVARTVIK, Jan

1966 *On voice in the English verb*. The Hague: Mouton.

TEEUW, A.

1978 *Leerboek bahasa Indonesia*. Groningen: Wolters-Noordhoff.

VERHAAR, John W.M.

1983 Ergativity, accusativity, and hierarchy. *Sophia Linguistica* 11:1-22.

APPENDIX

I am listing below a number of sentences having as contrasting pairs the meN- and the non-meN- forms, in order to give a better picture of the difference in terms of the verbal nuance between the two forms. They are quoted from Bur Rasuanto's novel *Tuyet* and Mochtar Lubis' novel *Harimau! Harimau!*

Verbs of perception

meN-	"Lihatlah ini!" kataku. *see this said-I* "SAYA MELIHAT!" kata pemilik kedai itu. (Bur Rasuanto, 42) *I see said owner shop that* *"Look at this!" I said.* *"I SAW it!" said the owner of the shop.*
non-meN-	"[....] Tapi kalau yang itu" - Thi menyentuhkan sikunya ke *but if* LIG. *that touch elbow-her to* tanganku - "orang yang berdiri dekat perempuan berbaju merah *arm-my person* LIG. *stand near woman have cloth red* muda di sana ... kaulihat?" *young in there you-see* "Ya, KULIHAT! Ada apa dia?" (Bur Rasuanto, 102) *yes I-see exist what ie* *"[....] But how about that person" - Thi nudged my arm - "the person standing close to the woman in pink over there ... do you see?"* *"Yes, I SEE him! What's the matter with him?"*

Verbs of consciousness

meN-	"Suratku itu barangkali memang tidak sampai ke tangan Herbert." *letter-my that perhaps indeed not reach to hand* "Mengapa KAU MENGIRA begitu?" tanyaku. (Bur Rasuanto, 81) *why you think so asked-I* *"Perhaps it's true that my letter didn't reach Herbert."* *"Why DO YOU THINK so?" I asked.*
	"[....] Dengan kata lain, kita sebenarnya sama." *with word other we in fact same* "Kesan saya sekarang justru sebaliknya", kataku. "Tadinya *impression my now* EMPH. *the reverse said-I before* SAYA memang MENGIRA begitu juga. [....]" (Bur Rasuanto, 147) *I indeed think so too* *"[....] In other words, we are of the same opinion."* *"I now think otherwise", I said. "Before I indeed THOUGHT that we were of the same opinion."*

Verbs of consciousness (cont'd)

meN-	"Oh, jadi, KALIAN MENYANGKA kalian [...] akan dapat *thus you*(pl.) *think you*(pl.) *will can* menangkap Wak Katok? [....]" (Mochtar Lubis, 205) *arrest* *"So, YOU THINK (come to the conclusion) that you can arrest Wak Katok just like that? [....]"*
non-meN-	"Ou, kalau itu aku tidak tertarik!" katanya. "KUKIRA tidak *if that I not interested said-her I-think no(t)* ada mahasiswa yang sungguh-sungguh tertarik." (Bur Rasuanto, 100) *exist student* LIG. *really interested* *"Oh, that I am not interested in!" she said. "I THINK there are no students who are really interested."*
	"Alimin, kau sadar apa yang akan kaulakukan ini?" *you aware what* LIG. *will you-do this* "Tentu saja aku sadar!" *of course I aware* "Tidak! KUKIRA kau tidak sadar! Kau ..." (Bur Rasuanto, 104) *no I-think you not aware you* *"Alimin, are you aware of what you are doing?"* *"Of course, I am!"* *"No. I THINK you're not aware of that. You ..."*
	"Ha-ha-ha", kata Wak Katok. "KALIAN SANGKA aku bodoh? [....]" *say* *you*(pl.) *think I stupid* (Mochtar Lubis, 188) *"Ha-ha-ha [laughing]", Wak Katok said. "DO YOU THINK I am stupid? [....]"*

Verbs of saying

meN-	"AKU cuma ingin MENYAMPAIKAN dua hal. Dan kau tak perlu *I only want tell two thing and you not need* berhenti bekerja untuk mendengarkannya. Satu: [....]" *stop work to listen to-it one* (Bur Rasuanto, 23) *"I'm only GOING TO TELL you these two things. And you don't need to stop doing what you're doing now while listening to me. First, [....]"*
non-meN-	"Aku tahu sekarang pun kau bisa pergi. Tapi kalau kau bisa *I know now* EMPH. *you can go but if you can* pergi sekarang, besok mungkin kau tak bisa ke mana-mana *go now tomorrow perhaps you not can to anywhere*

Verbs of saying (cont'd)

non-meN-	lagi .... Akan KUKATAKAN kepada papa kau sakit ..." *any more will I-tell to daddy you sick* (Bur Rasuanto, 141) *"I know that even now you can go. But if you go now, tomorrow you may not be able to go anywhere .... I'll TELL Dad that you're sick ..."*

Data reference

Ashadi Siregar, 1978 *Kugapai cintamu*. Jakarta: Gramedia.

Bur Rasuanto, 1978 *Tuyet: Kisah dari Negeri Perang*. Jakarta: Gramedia.

Mochtar Lubis, 1975 *Harimau! Harimau!* Jakarta: Pustaka Jaya.

Purwo, B.K. "The presence and absence of meN-: the Indonesian transitive verbs". In Geraghty, P., Carrington, L. and Wurm, S.A. editors, *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*.
C-93:159-170. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.159*

NUMBER IN BIAK: COUNTEREVIDENCE TO TWO ALLEGED LANGUAGE UNIVERSALS (A SUMMARY)

Hein Steinhauer

A slightly revised version of the FOCAL contribution with this title will be published elsewhere (Steinhauer 1985). A summary of its contents follows below.

The paper presents evidence against two language universals proposed by Joseph H. Greenberg in his article "Some universals of grammar with particular reference to the order of meaningful elements" (Greenberg 1966:73-113), i.e.

> "*Universal* 37. A language never has more gender categories in nonsingular numbers than in the singular" (p.95) and
>
> "*Universal* 45. If there are any gender distinctions in the plural of the pronoun, there are some gender distinctions in the singular also" (p.96).

Both these "universals" are posited as absolute laws, in contradistinction to most of the other universals proposed in the same article.

The language described in my paper is the Saui'as [sawi'as] dialect of Biak, which is spoken along the Straits of Sorendid'ori, which separates the islands of Biak and Supi'ori, to its south-east. My informant was Drs Johsz Mansoben from the village of Sorendiu'eri, who as an anthropologist has an excellent knowledge of many of the Biak speaking areas, including the dialectal differentiation found there.

Varieties of Biak have been described since more than a century ago, but - as is argued in this paper - most of the publications lack linguistic sophistication, while none of them goes into much detail.

The phonemic system of Saui'as Biak consists of the vowels /a, e, i, o, u/ and the consonants /b, p, v, f, m, d, s, n, r, ʔ/; word stress is phonemic. My paper discusses a number of problems connected with this phonemic system. Its main concern, however, is inflectional morphology.

The basic semantic categories relevant for the inflectional paradigms are illustrated in the chart of the independent personal pronouns (Chart 1).

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 171-173. *Pacific Linguistics*, C-93, 1986.

Chart 1

person / gender \ number		singular	dual	trial	plural
1	exclusive	/ai'a/	/nu/	/nʔo/	
	inclusive	-	/ʔu/	/ʔo/	
2		/'au/	/mu/	/mʔo/	
3	animate	/i/	/su/	/sʔo/	/si/
	inanimate				/na/

The same semantic categories in the same distribution are present in the subject agreement markers on (non-equative) predicates. Of these there are three sets. Which verb stem is connected with which set is only partly predictable. Variant forms are frequent for the second and third person singular and for the third person plural animate and inanimate.

The categories number and gender are also relevant for the demonstrative pronouns, which form a complicated, multidimensional set of oppositions. Chart 2 illustrates the core of this demonstrative system.

Chart 2

	singular	dual	trial	plural	
				anim.	inanim.
close to the speaker (*this*)	/in'e/	/suin'e/	/sʔoin'e/	/sin'e/	/nan'e/
relatively close to speaker and hearer (*that*)	/ii'i/ [iy'i]	/s'uii/ [s'uyi]	/sʔ'oii/ [sʔ'oyi]	/sii'i/ [siy'i]	/na'i/
relatively remote from speaker and hearer (*yonder*)	/iu'a/ [iw'a]	/s'uiua/ [s'uiwa]	/sʔ'oiua/ [sʔ'oiwa]	/siu'a/ [siw'a]	/nau'a/ [naw'a]
neutral (*the*)	/i/	/s'ui/	/sʔ'oi/	/si/	/na/

The morphemes constituting the forms of Chart 2 can be combined with other morphemes to indicate relative position of the designated entity with regard to various landmarks and locations defined by THE situation (usually the actual speech situation). At least some of the resulting demonstratives can be the base for further morphological extensions, which explicitly indicate that the designated entity is moving (for instance *towards*, *away from* or *past* the speaker). All the above forms can be explicitly marked for *givenness* (i.e. the hearer is assumed to know the designated entity). Many forms, finally, have variants required in specific syntactic positions (the forms of Chart 2 occur sentence finally, directly after a noun).

The prefixed possessor markers of the possessive pronouns show the same oppositions as the personal pronouns. The stem of the possessive pronouns is either /-van-, -an-, -ban-/ or /-ve-, -e-, -be-/; the former - marked - set again explicitly indicates *givenness* of the 'possessed' entity. The final parts of

the possessive pronouns, which are semantic and largely also formal copies of the demonstratives, refer to number, gender, relative position, etc. of that entity. The resulting forms have again syntactically conditioned variants.

In connection with the possessives the paper briefly discusses some irregular inflection types of the disappearing class of inalienable nouns.

The inflectional paradigms discussed in the paper as well as the system of independent personal pronouns show that the two Biak genders[1] are *only* distinguished in the plural. As there seem to be no other sections of the grammar where gender oppositions play a part, Biak presents counterevidence to the alleged language universals cited above.[2]

The paper concludes with a tentative explanation for this counter-intuitive state of affairs.

NOTES

1. It should be realised that the genders demarcate purely grammatical classes, which are not based on any synchronically self-evident bipartition of real-life phenomena, and that the terms "animate" and "inanimate" have only an approximative value.
2. A recent publication of the Percatakan Universitas Cenderawasih (O. Ramar et al. 1983) suggests that Wandamen contradicts the universals in question in a comparable way.

REFERENCES

GREENBERG, J.H., ed.

1966 *Universals of language*. 2nd edn. Cambridge, Massachusetts: MIT Press.

RAMAR, O., N. ONGKODHARMA and R. FLAMING

1983 *Vekavavo nana kavo Wondama, Percakapan-percakapan dalam bahasa Wandamen: Wandamen conversations*. Jayapura: Percetakan Universitas Cenderawasih.

STEINHAUER, H.

1985 Number in Biak: counterevidence to two alleged language universals. *Bijdragen tot de Taal-, Land- en Volkenkunde* 141/4.

ON THE SEMANTICS OF REDUPLICATION

Paz Buenaventura Naylor

As a morphological process, *reduplication* has been recognised in many languages of the world, as widely diverse typologically and genetically as English and Indonesian, Latin and Quechua, Sanskrit and Vietnamese - to name a few. As evidence of the occurrence of this phenomenon, the examples most often cited in general treatments have been onomatopaeic words in English and the perfect form of a small class of strong verbs in Classical Latin. The more detailed treatments, interestingly enough, have provided examples from languages outside of what Whorf referred to as SAE (Standard Average European). Yet it appears that grammarians and linguists alike have, in the majority of cases, not found it necessary to give a definition of *reduplication*. One can only assume that the word was perceived as self-explanatory. The base, *duplication*, means "doubling" or "copying", and this is exactly what is observed as having occurred. Why the *re-* was added, is a matter of conjecture. Perhaps it was an attempt to distinguish *reduplication* as a linguistic or technical term from "duplication" - a word that was not part of linguistic terminology and one which is interpretable in an 'everyday' or non-technical sense. Schachter and Otanes (1972), however, have adopted *duplication* in their treatment of this morphological process in Tagalog, thus imbuing this form with a technical linguistic interpretation.

A definition of *reduplication* as a linguistic term that refers to a particular morphological process is both necessary and useful toward the comprehension of the dynamics of reduplication. How, then, should we define *reduplication*?

All too often, and even amongst linguists themselves, *reduplication* has been taken to apply to all forms of repetition or 'duplication' of forms, thus rendering the word in its entirely literal sense. However, we do need to distinguish between the use of repetition as a rhetorical device and the use of repetition as a morphological device. Although these two uses of repetition do have certain semantic and pragmatic features in common, they differ in important ways in their functions and in the dynamics by which these functions are fulfilled.

Furthermore, such occurrences of repetition or duplication must be distinguished from that which results from the spurious or coincidental juxtaposition of two identical or very similar sound segments which, while in fact resulting in duplication of a sound segment bears no syntactic or morphological motivation for the duplicating process. We shall therefore exclude such instances of repetition or duplication as refrains in poetry or song and stylistic repetitions for emphasis or for mnemonic purposes as instances of *reduplication* in its linguistic sense. For example, repetition of the same sound segment (z) in a word like Harrises's or the repetition of the sound segment (n) in words like unenjoyable do not qualify as *reduplicated* forms either. Similarly, when the French

---

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 175-185. *Pacific Linguistics*, C-93, 1986.

first person plural pronoun nous occurs in "duplicate" as in Nous nous lavons les mains ici *We wash our hands here*, such duplication cannot be taken as an instance of *reduplication* either. (It would therefore be misguided to look upon *reduplication* as the antithesis of the "repeated morph constraint" (Menn and McWhinney 1984:519).

Before we go any further, let us take a look at some of the definitions that I did encounter. The entry for the word "reduplication" in Wyld's *Dictionary* (1961) - a very large compilation along etymological lines in the manner (but not the scale) of the OED - is the following: "the repetition of the first syllable of the word as in the Latin perfect; e.g., cecidi from caedere".

Bloomfield (1933:218) says: "Reduplication is an affix that consists of repeating part of the underlying form .... It may differ phonetically in some conventional way from the underlying word". As examples he gives, among others, the following: From Tagalog (su:sulat) *will write* from (sulat) *write*. From Fox, he gives the following paradigm:

(wa:pam :wa)	*he looks at him*
(wa:-wa:pam :wa)	*he examines him*
(wa.pa-wa:pam :wa)	*he keeps looking at him*

Bolinger (1978:116) says: "A third device in English is *reduplication*. The same morpheme is repeated, with or without modification: hush-hush, mish-mash, helter-skelter, fiddle-faddle".

Sapir (1921:76) says:

> Nothing is more natural than the prevalence of reduplication, in other words, the repetition of all or part of the radical element. The process is generally employed, with self-evident symbolism, to indicate such concepts as distribution, plurality, repetition, customary activity, increase of size, added intensity, continuance.

Schachter and Otanes (1972) do not give a definition of *reduplication* which as I have just pointed out, is *duplication* in their terminology. They describe two kinds of *duplication*: with subscript 1 (DUP_1), it refers to duplication of the first syllable of the root (referred to as *partial reduplication* elsewhere) and with a subscript 2 (DUP_2), it refers to reduplication of the first two syllables of the root (*full reduplication* elsewhere); since most roots are disyllabic, the duplication of the first two syllables is quite often full reduplication. Notice that like Bloomfield, they categorise reduplication as a form of affixation, referring to duplicated forms as verbal prefixes.

Too little is said in Bloomfield's and Bolinger's definitions: the repetition of either an unspecified morpheme (in Bolinger's definition) or of the underlying form (in Bloomfield's). It is Sapir's definition that brings out most of the attributes of *reduplication* and most of the meanings it conveys: repetition of the radical element, the notions of *process* and *symbolism*, and the expression of distinctions that, taken as a whole, may be said to be *aspectual* in nature.

Already the issue has been raised: is reduplication a form of affixation? If it is not, then *what* is it? As I have earlier mentioned, Bloomfield as well as Schachter and Otanes refer to *reduplication* as a form of affixation. Marantz (1984) likewise believes that *Tagalog reduplication* is affixation too. Others view *reduplication* as a morphological process distinct from affixation. Sapir (1921:62) in fact lists *reduplication* as one of six distinct grammatical processes, with *reduplication* kept quite distinct from affixation. It is not my intention

to argue on this matter in this paper which has an altogether different purpose. Nevertheless, since I am trying to establish clearer lines of definition of what *reduplication* is and is not, I shall state my position on the matter: *reduplication* is *not* affixation. There are many arguments that one could bring out to support this position but I shall give only one - the one that I believe to be the simplest and the best: it is simply that *affixation*, by definition, is the addition of an *affix* to a root; since *reduplication* is the repetition of part or all of a *root* - and here we are dealing with root components - how could we possibly say that we are dealing with *affixes* when we are *reduplicating* the root? It would be more accurate to say that with *reduplication* we have extended the root; but we certainly would not have found ourselves an *affix* by reduplicating the root.

While repetition in itself does have rhetorical value and pragmatic function, we need to define *reduplication as a morphological process* in narrower terms. In this sense, any repetition of any morpheme does not necessarily constitute *reduplication*. Chance juxtaposition of two identical or very similar phonological segments does not constitute *reduplication* either. For *reduplication* to be said to occur in this specifically linguistic sense, the duplication *must be motivated* by language-specific morphosyntactic rules. Both root and duplicate segment must constitute the same higher morphological unit, i.e. they must be constituends of the same constituent.

Reduplication is a morphological process by which all or part of the root is repeated to express a variety of syntactic or semantic distinctions. The duplicate may be subject to phonetic modifications based on language-specific rules. It appears to occur in one form or another in most languages. In some languages, however, *reduplication* exists mainly as fossilised lexical items and is of very limited productivity - if it can be said to be productive at all. On the other hand, in certain other languages, *reduplication* occurs as a highly productive morphological process that plays an important role in the grammar and in the lexicon.

As Sapir said: "Even in English it (reduplication) is not unknown, though it is not generally accounted one of the typical formative devices of our language" (1921:76). Although he gives examples from English such as "a big, big man" and "sing-song, wishy-washy, roly-poly", etc. he nonetheless concludes that "... it can hardly be said that the duplicative process is of a distinctively grammatical significance in English" and "we must turn to other languages for illustration" (1921:77) - as indeed we shall. The fact that it does not form a significant part of the morphological system of English is reflected in the fact that in such a widely used textbook for introductory courses in linguistics as Fromkin and Rodman's (1978), they did not see fit to include *reduplication* in the chapter on morphology (although among the exercises at the end of the chapter on morphology, there was one on Samoan in which *partial reduplication* marked the plural form of the verb). Had this chapter on morphology been part of a book on the teaching of English, it would hardly matter that *reduplication* was not included in the discussion of morphology. On the other hand, it would be impossible to teach Tagalog without explaining the function and meaning of *reduplication* and its role in the grammatical and lexical processes of Tagalog. In Tagalog, *reduplication* is part of the "core grammar" whereas in English it is not.

In view of the preceding discussion, we can say that *reduplication* as a regular and productive morphological process does not occur in languages such as English or what Whorf referred to as Standard Average European. On the other hand, those of us who have worked with languages in which *reduplication* is part

of the core grammar know what a productive - and fascinating - morphological process *reduplication* is. Among the articles dealing with *reduplication* in Austronesian languages, we have Alieva on Indonesian (and apparently on Tagalog as well but her work has remained inaccessible to me since it is written in Russian), Flora on Palauan (1978), Harrison on Micronesian languages (1974), Lopez (1950) and Blake (1917), and Carrier (1979) on Tagalog. On the whole, however, the approach in the treatment of *reduplication* has thus far been taxonomic or formal-derivational. I have not come across any attempts to seek out the underlying common denominator that characterises the various and apparently divergent functions and meanings of reduplication and unifies them into a well-defined and coherent system. This paper is one such attempt.

In this paper, I propose the view that *reduplication is a semantic system based on aspectual contrasts mediated by the iconicity of the duplicated form that is its morphological realisation.*

That *reduplication* is iconic has been pointed out by Haiman (1980 and 1983) and Diffloth (1982). Such iconicity is in fact self-evident. *Reduplication*, as it doubles all or part of the root, serves as iconic representation of imperfectivity; duplication doubles the point of reference, thus embodying iconically, not just symbolically, the multipunctual perspective expressed by aspectual imperfectivity. The form is a *concretisation* of the duplicative meaning: the form is an *icon* of the concept. The word *is* the thing. It is semantically transparent.

It is easy enough to see how many of the meanings of reduplicated forms such as repetition, plurality, intensification and the like, could be drawn from the "doubling" of the point of reference that results from the doubling of the form. The additive force of iconic doubling is obvious. What is less obvious and almost in apparent contradiction, is the detractive meaning that certain reduplicated forms convey, such as the diminutive or disdainful attitude. Unless we take the analysis a step further, a simple "iconic doubling" explanation will not explain how an additive process such as doubling can convey detractive or subtractive meaning. Furthermore, how could such an explanation account for the function of *reduplication* as marker of the imperfective in the Tagalog verbal aspect paradigm? Neither would it account for the meaning of likeness or facsimile that certain reduplicated forms convey. It is clear from an examination of the various meanings that *reduplication* conveys that the kind of iconicity that *reduplication* as a morphological process presents is *relational*. Since *reduplication* results in the doubling of a segment, it also results in the presence of more than one point of reference and a change of perspective from absolute to relative. The view from a relative perspective can go in either direction. Once we are able to assume a relative perspective, we can then readily account for the fact that *reduplication* conveys both additive as well as subtractive meanings, comparative as well as other relative meanings. In its function as marker of the imperfective in Tagalog, iconic doubling, as it creates another point of reference results in the multipunctuality that characterises verbal imperfectivity.

Let us now examine examples of reduplication in Tagalog that should illustrate the points that I have just brought out in the preceding paragraphs.

1. The verbal aspect paradigm

aral *study*

NOT BEGUN		BEGUN	
INFINITIVE	CONTEMPLATIVE	INCOMPLETIVE	COMPLETIVE
mag-aral	mag-aaral	nag-aaral	nag-aral
to study	*will study*	*studies*	*studied*
		is studying	*has studied*
		was studying	*had studied*

Notice that the reduplicated form crosses the boundary between the *begun* and *not begun* aspects (*realis versus irrealis*). It is clear that within these categories, there is the further distinction of imperfective and perfective with the imperfective marked by *partial reduplication* of the root.

The nominal form of verbs also **has** the partially reduplicated root plus the prefix pag- (which in turn can serve as the base for "re-verbalisation"). Hence, the nominal form of mag-aral is pag-aaral *the act of studying.*

There is a form in Tagalog that parallels pag-aaral but whose meaning is verbal imperfective aspect. It is what Schachter and Otanes (1972:371) have labelled the "recent perfective", e.g. kaaaral (the partially reduplicated root plus the nominal prefix ka-). It means *just finished studying* or *just studied.* The meaning is obviously verbal but the form is nominal and is subject to noun syntax. This distinction, on the surface, may seem irrelevant to the subject at hand but as will soon be clear, it is crucial to the analysis of this aspectual form, the notion of aspect and its scope of applicability in general and to the semantics of reduplication in particular.

The following sentences are in the recently-completed aspect:

(1) Kaaaral ko lang.
just finished studying my only
I have just finished studying.

(2) Katatapos lang ng programa.
just finished only of the program
The program just ended.

Literally, these sentences mean *My (act of) just finished studying* and *The program's just having been ended*, respectively. Roots without any marking such as affixes, particles, or contrastive stress are neutral as to grammatical function. Syntactically, they are neither noun, verb, nor adjective; they are purely semantic concepts and they remain so until marked for a specific syntactic function. Since *reduplication* applies to roots (with very few exceptions at least in Tagalog), and since, as the examples that are given throughout this paper will show, the semantic distinctions that the various forms of reduplication convey are aspectual in nature, then we can say that aspectual distinctions apply indiscriminately of part-of-speech function of the duplicated morpheme. In other words, imperfective (or perfective) aspect is a property of verbs as well as nouns and adjectives, provided that the lexical content of the word includes such features as duration, phases, continuity, punctuality, etc. that are associated with aspect, which, in the final analysis, refers to the perspective in which the speaker wishes to present what he is talking about. To return to the so-called "recent perfective" in Tagalog, the lexical content of the root aral is presented as aspectually imperfective, but with the prefix ka-, the whole word, kaaaral, refers to the completion of a durative (imperfective) act. At the same

time, it conveys relative time reference, i.e. the time of completion of the situation (recent) relative to the moment of speaking.

The so-called "recent perfective" form in Tagalog is, therefore, a *perfect* form, not a *perfective*. It marks the *completion* of an *imperfective situation*.

In Tagalog, *reduplication* is not used to pluralise nouns as it is in Indonesian, for example. It is however used to form the (optional) plural of modifiers (adjectives/adverbs) and verbs as in the following examples.

2. Pluraliser

Adjective - magaganda (from maganda *beautiful*)
Participial adjective - dala-dala (from dala *carried/borne by*)
Verb - nangag-aral (from nag-aral as above)
magkita-kita (from magkita *to meet each other*)
to meet one another (more than two people)

3. Plural plus randomness

magkawalawala (from mawala *get lost*)
get lost all over the place

magkanghuhulog (from mahulog *get dropped*)
fall all over the place

These two verbs and others like it can only be used in the plural.

Reduplication also conveys *moderative* aspect in verbs of the form used in the following examples.

4. Moderative aspect

> Moderative verbs express activities performed perfunctorily, occasionally, at random, etc. They are often equivalent to English verbs accompanied by such phrases as 'a little', 'a bit', 'now and then', 'here and there', etc. One of their most common uses is to soften a request ....
> (Schachter and Otanes 1972:340)

(a) Magwalis-walis ka nga ng bakuran.
Please SWEEP the yard A LITTLE.
(cf. Magwalis ka nga ng bakuran *Please sweep the yard*)

(b) Laki-lakihan mo ang putol.
MAKE the cuts (slices) A LITTLE LARGER.

(c) Dagda-dagdagan mo ng tubig ang linaga.
ADD water to the stew EVERY NOW AND THEN.

5. Moderative plus randomness

Maglalakadlakad muna kami habang wala pa sila.
We'll walk around a bit while they are not yet here.

Other categories of meaning that *reduplication* conveys are the following.

6. Facsimile

bahay-bahayan *play-house* (from bahay *house*)[1]
anak-anakan *foster child* (from anak *child/offspring*)
tao-tauhan *figurine/toy man* (from tao *person*)

(a) Nagpilaypilayan si Lito.
Lito (pretended to be lame).

(b) Naglalaro ng bahay-bahayan ang mga bata.
The children are (playing house).

7. Detractive/subtractive

(a) Ano bang libe-liberal? Pare-parehong lahat iyan.
What do you mean "liberal"? They are (all the same).

(b) Tama na. Wala nang disku-diskusyon.
That's enough. No more/no further discussion of any sort.

(c) Naku! May kasti-kastila pa siya!
Ugh! She's putting on airs using Spanish! (lit. *She even has that (fancy) Spanish!*)

(d) Huwag ka nang sumagotsagot. Talagang mali ka.
Stop your (pointless) answers (lit. *answering). You're definitely wrong.*

Adjectives: magandaganda *somewhat good looking* (from maganda *beautiful*)
malaki-laki *kind of big* (from malaki *big*)[2]

Adverbs: madali-dali *rather easy* (from madali *easy*)
mabilis-bilis *rather fast* (from mabilis *fast*)

8. Intensification

Adjectives: magandang maganda *very beautiful*
malaking malaki *very big*

Adverbs: madaling- madali *very easy*
mabilis na mabilis *very fast*

(a) Pagkagandaganda!
What great beauty!

(b) Ang gandaganda!
What a very beautiful thing!

(c) Pagod na pagod na si Maria kasi aral nang aral.
Maria is (very tired) now because she (studies all the time).

The second reduplication, aral nang aral is repetitive.

9. Repetitiveness

(a) Wala kaming ginawa kundi kumain nang kumain.
We did nothing but (eat all the time) (lit. *eat and eat*).

(b) Pinaghahalikan ako sa bibig, pinaghihipo kung saan saan sa katawan.
(He) repeatedly kissed me on the mouth, (he) repeatedly fondled (me) everywhere on my body. (From an interview with a rape victim, *Liwayway* magazine.)

This sentence also connotes randomness but it is a function of the base, not of the reduplication.

10. Distributive

isa-isa (isa *one*) *one by one*
bahay-bahay (bahay *house*) *every house*

araw-araw (araw *day*) *every day*
tig-sisingkwenta (sinkwenta *fifty*) *fifty cents apiece*

From the foregoing examples, it should be fairly obvious that the various categories of meaning that *reduplication* conveys denote some form of aspectual distinction. In fact, the examples of *reduplication* from a wide variety of languages that I have seen had this in common. The kinds of meaning that one could expect reduplicated forms to convey - and did in fact convey - had a familiar ring in that they involved aspectual distinction of one sort or another.

To recapitulate, the core of the semantic system within which the various meanings of *reduplication* are realised is *multipunctual perspective or imperfectivity*. This is not to say that all the various meanings that derive from *reduplication* are "reduced" to a single imperfective meaning. What we are trying to say is that aspectual imperfectivity is the underlying semantic from which the various subcategorial meanings derive their force. Perhaps the following diagram will help to explain the explanatory hypothesis presented in this paper.

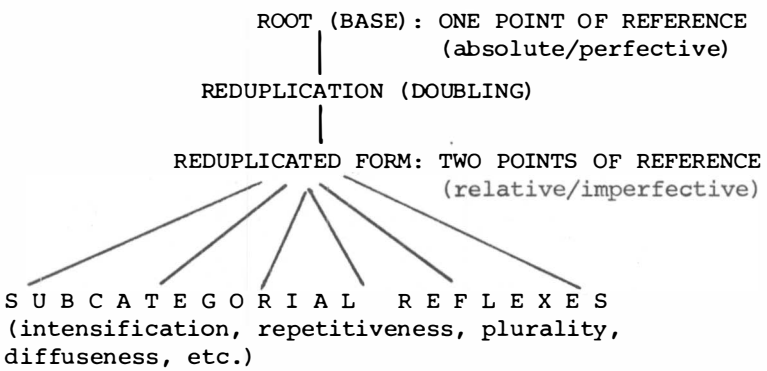


Figure 1

The diagram schematises what I have earlier pointed out. When *reduplication* occurs, the doubling of the root segment provides iconically a double point of reference, moving into the multipunctual perspective that characterises imperfectivity. Aspectual imperfectivity is a *necessary correlate* of *reduplication*. At the lower nodes, the lexical content of the root and other auxiliary morphology contribute to the individuated meanings - what we may consider the subcategorial reflexes of the underlying semantic - the imperfective.

Since aspect reflects perspective, then in certain uses of *reduplication* in Tagalog, it becomes a vehicle for the expression of the speaker's feelings (e.g. admiration or disdain) towards a person, thing, or situation. The range of meaning of *reduplication* in Tagalog may therefore be said to go from the realisation of contrast in the aspect paradigm (verbal morphology) to the expression of feelings that in themselves are not conveyed by the lexical content of the reduplicated root.

In conclusion, we wish to say that, from the analysis of reduplicated forms we found in extensive samples of written and spoken text, the hypothesis that emerges is that imperfectivity of aspectual perspective is the underlying semantic-pragmatic principle that serves as the common denominator uniting the various and sometimes seemingly unrelated meanings of *reduplication* as a morphological process.

It is interesting to note that *reduplication* as a productive morphological process is part of the linguistic system of languages that may be characterised as predominantly pragmatically based in opposition to languages as the Western European which are classifiable as predominantly syntactically based (after Givón 1980). This points to a pragmatics that goes beyond morphology and surface syntax. This may also point to *reduplication* as a typological feature. The "multifunctionality" of *reduplication* associated with what I have labelled "subcategorial (semantic) reflexes" in the diagram above (Figure 1), has been taken to be a diagnostic feature of a language type whose definition appears to be the same type that Givón refers to as predominantly pragmatically based, and which appears to be what Austronesian languages generally are.

NOTES

1. George B. Milner pointed out a parallel example in Fijian: waqa *boat*, waqa-waqa *spiritual medium* (personal communication).
2. cf. Fijian lailai *lesser quantity* (diminutives are also reduplicated in Fijian) as G.B. Milner has informed me (personal communication).

REFERENCES

ABBI, Avita

1985 The concept of simultaneity and iteration in Indian languages. Manuscript for *SALA* 7, Ann Arbor, Michigan.

BLAKE, Frank L.

1917 Reduplication in Tagalog. *American Journal of Philosophy* 38/4:425-431.

BLOOMFIELD, Leonard

1933 *Language*. New York: Holt, Winston and Rinehart.

BOLINGER, Dwight

1975 *Aspects of language*. New York: Harcourt, Brace, Jovanovich.

BAT-EL, Outi

1984 Reduplication in Hebrew: a case of *re* reduplication. Abstract, LSA Meeting Handbook.

BERLIN, B.

1963 Some semantic features of reduplication in Tzeltal. *IJAL* 29:211-218.

BOURCIEZ, E.

1946 *Éléments de linguistique romane*. Paris: Librairie Klincksieck.

CARRIER, J.L.

1979 The interaction of morphological and phonological rules in Tagalog. Doctoral dissertation, MIT, Cambridge, Mass.

COMRIE, Bernard

1976 *Aspect*. Cambridge: Cambridge University Press.

DIFFLOTH, Gérard

1980 Expressives and iconicity: do we need an aesthetic component? Paper given in Ann Arbor, Michigan.

EVERETT, Dan and Lucy SEKI

1984 Deletion, reduplication, and CV skeleta in Kamaiura. LSA Meeting Handbook.

ELCOCK, W.D.

1960 *The Romance languages*. London: Faber and Faber.

FROMKIN, V. and R. RODMAN

1978 *An introduction to language*. New York: Holt, Rinehart and Winston.

FLORA, JoAnn

1978 Reduplication in Palauan. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, Fascicle 1:617-655. *PL*, C-61.

GIVÓN, T.

1980 *On understanding grammar*. New York: Academic Press.

HAIMAN, J.

1980 The iconicity of grammar: isomorphism and motivation. *Language* 56/3:515-540.

1983 Iconic and economic motivation. *Language* 59/3:781-819.

HARRISON, S.P.

1974 Reduplication in Micronesian languages. *Oceanic Linguistics* 12: 407-454.

HEALEY, P.M.

1960 *An Agta grammar*. Manila: SIL, Philippine Branch.

LOPEZ, C.

1950 Reduplication in Tagalog. *Bijdragen tot de Taal-, Land-, en Volkenkunde* 2:151-311.

LYONS, J.

1977 *Semantics*. vols 1 and 2. Cambridge: Cambridge University Press.

MARANTZ, A.

1982 Re reduplication. *Linguistic Inquiry* 13/3:435-482.

1984 Tagalog reduplication is affixation too. LSA Meeting Handbook.

MARKEY, T.L.

1984 Multifunctionality as a concept and definer of lectal typologies. Paper presented at the Fifth International Conference on Methods in Dialectology, University of Victoria, B.C., Canada.

MENN, L. and B. McWHINNEY

1984 The repeated morph constraint: toward an explanation. *Language* 60/3: 519-541.

NAYLOR, P.B.

1978 Expressions of irrealis in some Philippine languages. *University of Michigan Publications in Linguistics* 2/4:131-143.

NHAN, N.T.

1984 Reduplication, 'interlocking' phrases, and cocompounding in Vietnamese. LSA Meeting Handbook.

QUINE, W.N.O.

1960 *Word and object*. Cambridge, Mass.: MIT Press.

ROSEN, Joan

1972 Reduplication in Indonesian. MS.

SAPIR, E.

1921 *Language*. New York: Harcourt, Brace, Jovanovich.

SCHACHTER, P. and F. OTANES

1972 *Tagalog reference grammar*. Los Angeles: University of California Press.

ULRICH, C.

1984 Reduplication as infixation in Puma and Papago. LSA Meeting Handbook.

WHORF, B.L.

1956 The punctual and segmentative aspect of verbs in Hopi. and The relation of habitual thought and behavior in language. In J.B. Caroll, ed. *Language, thought, and reality: selected writings of Benjamin Lee Whorf*. Cambridge, Mass.: MIT Press.

WYLD, H.C.

1961 *The universal dictionary of the English language*. London: Routledge and Kegan Paul.

PALAUAN AS A VOS LANGUAGE

Carol Georgopoulos

1. INTRODUCTION

In this paper I would like to outline some arguments as to the basic word order of sentences in the language of Palau, spoken in the Western Caroline Islands. I hope to show that these considerations will both contribute to the description of Palauan grammar itself, and add to the evidence demonstrating Palauan's affinity to other languages in its language family. Basic word order is related to the actual word order of spoken sentences in a systematic way, and I will describe this relation in terms of the productive sentence types of Palauan.

Although this paper at some points refers to Josephs' *Palauan reference grammar* (1975), the major work on Palauan, my primary intention is not to criticise that work, but rather to provide an independent description of Palauan grammar. My analysis departs from that of Josephs especially in the matter of the order of constituents. Josephs represents the basic word order of Palauan as subject-verb-object (SVO), and proposes a set of rules that account for surface changes in this order. Waters (1979) argues that Palauan is VOS and demonstrates that Palauan sentences in which a noun phrase precedes the verb are produced by topicalisation; he shows, for example, that a NP in this position may bear the grammatical relation of direct object or oblique object as well as subject. This paper expands the hypothesis that basic Palauan word order is VOS.

2. SOME PALAUAN SENTENCE TYPES

Although word order in Palauan is flexible, it is by no means 'free'. A transitive sentence with the order V-NP-NP has only the interpretation on which the first NP is object and the second is subject. There is no case marking of NPs, so this order is crucial to identifying grammatical role; see the examples in (1) (abbreviations used in the glosses are explained in note 1).[1]

(1)a. toltoir a katu a beap
3p-*chase* *cat* *mouse*
The mice are chasing the cats
**The cats are chasing the mice*

b. ngkiltmeklii a ulaol a Peter
3s-*clean* *floor*
Peter cleaned the floor

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics,* 187-198. *Pacific Linguistics*, C-93, 1986.

c. tomes a 'arm a re'ad er a siabal
3p-*see* *animal men* P *Japan*
The Japanese are looking at the animals
**The animals are looking at the Japanese*

'Scrambling', therefore, as in Walpiri, Tagalog, or Latin, is not possible in Palauan. Subject agreement is prefixed to the verb:

(1)d. te-'illebedii a bilis a rengalek
3p-*hit* *dog* *children*
The kids hit the dog

Simple sentences frequently begin with a noun phrase, as in the examples in (2) (the grammatical relation of the NP is given in parentheses):

(2)a. a 'ermek a soal el melim (subject)
dog-my R-*want* L R-*drink*
My dog wants to drink

b. a blai a lesilsebii a se'elik (direct object)
house IR-3-*burn* *friend-my*
My friend burned down the house (Josephs 401)

c. a Naomi a le'ilitii a 'ole'esel a John (possessor)
IR-3-*throw* *pencil-her*
John threw away Naomi's pencil

d. a blil a Irene a ngar ngii a bung er a medal (oblique object)
house-her *be* *it* *flower* P *front-its*
There are flowers in front of Irene's house

e. a bdelula'ang a leble er ngii a bilas (prepositional object)
pier IR-3-*come* P *it* *boat*
The boat came to the pier

It is important to note that, in the sentences in (2), linear order does not reveal grammatical role. In addition, in (2b) through (2e), the verb agrees with the subject, although the subject is not in initial position; the sentences are transitive, and the initial NP may be a possessor or a locative as well as a direct object. The facts of (2) are consistent with the topicalisation hypothesis: topicalisation is not limited to certain NPs, as passive is, and in topicalisation phrases are reordered without affecting transitivity or grammatical role.[2]

Now consider the sentence type which Josephs (1975) claims to be passive. In Josephs' definition of the passive, the verb has irrealis morphology ("hypothetical", in Josephs' terms), and a pronoun copy of the passive subject may appear in object position whenever that position is marked by a preposition.[3] The examples in (2b), (2c) and (2e) fit this description, as does the example taken from Josephs, seen in (3):

(3) a rengalek a longelebed er tir a sensei
children IR-3-Im-*hit* P *them* *teacher*
Josephs: *The children are being hit by the teacher*

A problem that immediately arises for this analysis is that the verb in (3) is transitive. Transitivity in imperfective verbs is indicated by an infixed

'imperfective marker'; in (3) this morpheme has the form -ng-.[4] The transitive morphology of (3) contradicts what we would normally expect of a passive. In addition to this, there are other constructions in Palauan that are more naturally analysed as passives. Compare to (3) the sentences in (4):

(4)a. a suebek el 'arm a uleboes
bird L *animal shot*
The bird got shot

b. a buk er ngii a dimlak leme'uiu
book P *him* Neg *read*
His book was not read

c. a rengalek a me'elebed
children *hit*
The kids are being beaten

(I do not gloss the mood of the verb in these examples so as not to force the analysis.) These sentences have intransitive morphology: they lack either of the transitivising morphemes of Palauan.[5] They also have passive meaning, in that Palauan speakers use them in the same contexts that would require the passive in English: the subject is theme or patient, the agent is not normally expressed, and so on. The subjects in such sentences often, though not obligatorily, appear before the verb.

I would like to explore a quite different account of the difference between sentences like (3) and those in (4). In order to do so, however, a brief foray is necessary into the more general background of Palauan grammar.

2.1 Overview of the Palauan verb phrase

A transitive verb in Palauan has either imperfective or perfective morphology; only transitive verbs are inflected for aspect. Verbs of either aspect carry subject agreement (which is prefixed). In addition, a perfective verb carries direct object agreement (which is suffixed). An imperfective verb does not agree with its object, but rather marks that object with Palauan's universal preposition, written er ([ər] or [ɾ̩]). Marking of a direct object in either of these ways occurs only when that object is *human*, or when it is *specific* in some sense defined by the language. These facts are illustrated in (5):

(5)a. ngulemeng' er a 'o'il a bilis
R-3s-Im-*bite* P *foot-his dog*
The dog was biting its own foot

b. akuldengesterir a resensei er ngak
R-1s-Pf-*honour*-3p *teachers* P *me*
I praised my teachers

c. ngomekeroul a bung a Rose er a sersel
R-3s-Im-*grow* *flower* P *garden-her*
Rose is growing flowers in her garden

d. te'illebed a bilis a rengalek
R-3p-Pf-*hit* *dog* *children*
The kids hit some dogs

(5a) contains an imperfective verb, whose specific object 'o'il is marked with er; (5b) contains a perfective verb which carries the third person plural object morpheme, -terir, referring to resensei (human plural); the direct object in (5c), bung, has no marker (no er) because it is plural, nonhuman (plurality of nonhuman objects is not marked morphologically); similarly, the absence of object agreement on the perfective verb in (5d) indicates that the object, bilis, is plural and/or nonhuman.

These details, admittedly somewhat complex, underly the discussion of the next section.

3. TOPICALISATION

A noun phrase of any grammatical relation may be in clause-initial position (cf. the sentences in (2)). Waters (1979), in arguments based on the facts of causatives, equi (both subject- and object-controlled), comitatives, and other structures, shows that such an NP frequently lacks the properties of a subject. I will not describe the pragmatics of NP preposing here. There are, however, two important syntactic constraints on such a construction. I will take each of these constraints in turn.

The first affects constructions in which the initial NP is a direct object, and stems from the disparity in object marking I have just described. When the object of a perfective verb is in first position, there is a gap in the object slot following the verb. Recall that person and number information relevant to this object, and therefore relevant to the gap, is present in the object agreement morpheme suffixed to the verb.[6] This is illustrated in (2b), repeated in (6), where the gap is indicated by a line, and object agreement is underlined.

(6) a blai a lesilsebii ___ a se'elik
house IR-3-Pf-*burn*-3s *friend-my*
My friend burned down the house

In contrast, fronting of the object of an imperfective verb is accompanied by the obligatory appearance of a pronoun following the verb. This pronoun has person and number features matching those of the clause-initial NP. Note that this position is marked by a preposition, which does not carry agreement, so the person and number information about the fronted NP is provided by the pronoun. These facts are seen in (7). Compare object position in (6) and (7):

(7) a bilis a lomes er ngii a ngalek
dog IR-3-Im-*see* P *it* *child*
The child is looking at the dog

Within simple sentences like (6) and (7), the presence or absence of agreement (or a pronoun) does not appear to be crucial in linking the preposed NP with its origin site. But in complex sentences the same constraint holds, helping to identify the position corresponding to the distant NP:

(8) a bilas a lodengelii a rire'e'li ____ a tonari er kemam
boat IR-3-Pf-*know*-3s R-Pf-*steal*-3s *neighbour* P *us*
Our neighbour knows who stole the boat

To sum up to this point, inflection for aspect interacts with the verb's morphology with respect to the direct object, and this morphology alternates with the form in object position when the object is fronted. There is a zero form in this position when the verb carries object agreement, and a retained pronoun otherwise. In the examples in (4), on the other hand, verb morphology is intransitive, expression of agents is severely constrained, and there is no pronominal trace, as Josephs calls it, of the (derived) subject. The facts presented so far argue that (3) is a topicalised, transitive sentence, and those in (4) are passive.

The second constraint on the constructions that are analysed here as topicalisation has to do with mood morphology. There are both semantic and syntactic factors determining the mood of a clause. The semantic factors are negation, conditionals, wishes, and the like, which I will not illustrate here. But there are also strictly syntactic factors. Briefly, topicalisation of a subject is accompanied by realis verb morphology, and topicalisation of a nonsubject correlates with irrealis morphology. This can be seen in many of the examples above. The examples in (9) contrast in just this feature: realis morphology accompanies subject topicalisation and irrealis morphology a nonsubject topic.

(9)a. a rengalek a rirellii a present ___ el mo er a sensei
children R-Pf-*do*-3s L *go* P *teacher*
The children made a present for the teacher

b. a present a llirellii ___ a rengalek el mo er a sensei
IR-Pf-*do*-3s
The children made a present for the teacher

To return now to the example from Josephs, example (3), we see that the verb is transitive; that a pronoun copy of the object follows the imperfective verb, this pronoun having the features of rengalek *the children*; and that the verb is realis. Further, no semantic factor requiring irrealis mood is present. The facts are analogous to those in (7) and in (9b). I conclude, therefore, that the sentence (3) is a topicalisation construction rather than a passive, and the correct gloss is *The teacher is hitting the children*. Note especially that the subject follows the object in (3), as in (7), and (9b).[7]

4. OTHER 'EXTRACTION' CONSTRUCTIONS

Having worked through an explanation of the occurrence of retained pronouns and of mood alternations in transitive sentences, we may now ask whether other constructions provide support for the conclusion that these sentences are cases of topicalisation. Generative grammarians have found that topicalisation structures have many properties in common with certain others, such as WH-questions and relativisation (see, for instance, Chomsky 1977). The facts of Palauan grammar provide strong support for this finding: WH-questions and relativisations have the properties that are described in section 3, as the discussion below will show.

4.1 WH-questions

A WH-phrase may be in situ, or it may be preposed. Both types are seen in (10):

(10)a. kemil'erar tiang el mo er te'ang
R-2s-Pf-*buy*-3s Dem L *go* P *who?*
Who did you buy this for?

b. kemesuub a ngera el tekoi er a 'el'ang
R-2s-Im-*study* *what?* L *language* P *now*
What language are you studying now?

c. ngngera a lesilsebii a se'elil
what? IR-3-Pf-*burn*-3s *friend-his*
What did his friend burn?

Initial position of a WH-phrase is accompanied by the mood alternations described above, and a retained pronoun may appear in the argument position corresponding to the WH-phrase, as illustrated in (11):

(11)a. ngte'a a 'omulme'ar tiang el mo er ngii
who IR-2-Im-*buy* Dem L *go* P *him/her*
Who did you buy this for?

b. ngte'a a kileldii a sub
who R-*heat*-3s *soup*
Who heated up the soup?

c. ngngera el rum a lulngetmokl er ngii a Willy
what L *room* IR-3-Im-*clean* P *it*
Which room did Willy clean up?

Compare (11a) to (10a). This pair illustrates the same contrasts that we saw above for NP-initial declaratives: a basically realis question ((10a) shows irrealis morphology when a nonsubject NP is fronted (11a)), and a pronoun is 'left' in the position of the fronted WH-phrase.

4.2 Relativisation

A relative clause follows the head noun, and is joined to it by the 'linker' morpheme, el; there are no relative pronouns in Palauan. Relativised subjects do not appear. Relativised nonsubjects may or may not appear in pronoun form, depending on the presence or absence of agreement or of er associated with the relative NP. Examples of relative clauses are seen (in square brackets) in (12).

(12)a. akmedengelii a 'ad [el mil'erar tia el buk ___]
R-1s-Pf-*know*-3s *man* L R-Pf-*buy*-3s Dem L *book*
I know the person who bought that book

b. a babier [el lil'esii ___ a sensei] a mildul er a bulis
letter L IR-3-Pf-*write*-3s *teacher* Ps-*burn* P *police*
The letter the teacher wrote was burned by the police

c. a 'ad [el kul'erar a ngikel er ngii] a demak
man L IR-1s-Pf-*buy*-3s *fish* P *him* *father-my*
The man I bought the fish from is my father

It can also be seen in the examples in (12) that the mood of the verb is realis when the subject is relativised (as in (12a)), and irrealis when a nonsubject is relativised (as in (12b) and (12c)).[8]

The appearance of retained pronouns following a preposition, and the connection between grammatical relation and the mood expressed by the verb, therefore, are properties shared by relative clauses, WH-questions, and the sentences in (6) through (9). This sharing of properties demonstrates that all these constructions are of the same syntactic type. In terms of generative grammar, all three constructions are produced by WH-movement; since they are generated by the same rule, WH-movement structures have the same essential properties. Finally, the shared properties of topicalisation, WH-questions, and relativisation structures provides support for my claim that the NPs in preverbal position in examples (6) through (9) are in a position of focus,[9] an interpretation confirmed by the intuitions of Palauan speakers.

5. ORDER IN DEPENDENT CLAUSES

It has been observed, first by Ross (1967), that reordering of phrases most commonly takes place in main clauses, and that embedded clauses are better indicators of a language's basic word order. Certainly study of a variety of dependent clause types is necessary before arriving at conclusions about basic word order. Although Palauan embedded clauses may be headed by a topic (subject or nonsubject) or by a WH-phrase, they typically have VOS or verb-initial order. Examples of this are seen in (13) (dependent clauses are bracketed):

(13) a. a babier [el lil'esii a sensei] a mildul er a bulis (=(12c))
letter L IR-*write* *teacher* Ps-*burn* P *police*
The letter the teacher wrote was burned by the police

b. akulmes er a buk [el lulme'ar er ngii a 'ad er a Sie]
R-1s-*see* P *book* L IR-Im-*buy* P *it* *man* P
I was looking at the book that the man bought from Sie

c. akmilnguiu er a buk [er sei er a lemei a Moses]
R-1s-Im-*read* P *book* P *when* P IR-*come*
I was reading a book when Moses came

d. 'omomdasu [e ngmilskak a ngera a loseb er a skuul]
IR-2-*think* Prt R-3s-*give-me* *what* P *school*
What do you think Joseph gave me at school?

e. ngte'ang a lilsang a Peter [el lulengelebed er ngii a Mary]
who IR-*see* L IR-Im-*hit* P *him/her*
Who did Peter see Mary hit?

f. a ltelbang a omdasu [el kmo ngmo me'ar er a bilas a ngelekel a se'elik]
think *that* R-*go* R-*buy* P *boat* *child-his* *friend-my*
Itelbang thinks that my friend's boy will buy the boat

6. JOSEPHS' REORDERING RULES

In the analysis in Josephs' *Grammar*, subject-final sentences are derived by a rule of Subject-Shifting: the subject which would occur naturally before the verb is postposed to the end of the sentence. Although there is no morphological evidence that the subject originates before the verb (or, for that matter, that it originates to the right of the verb), the following points may be made. First, no motivation of the subject-shifted construction is given, and

we may ask why it applies only to subjects. In the VOS hypothesis, on the other hand, subjects are naturally sentence final. These are not just two sides of the same coin, as the VOS hypothesis accounts for SVO orders with no special statement about subjects; the topicalisation rule is needed independently for other NP-first constructions.

Second (this is not directly a criticism of Josephs, in that he does not describe constraints on his subject-shifted construction), subjects in Palauan are ordered rigidly with respect to objects (see example (1)), but may occur either before or after oblique phrases. This can be seen in (13b), (13f), and in (14):

(14) ng'illebedii ^{x} a buik a sensei er a skuul ↓ er a 'elii ↓ el oba a kerrekar ↓
R-3s-Pf-*hit*-3s *boy* *teacher* P *school* P *yesterday* L *use* *wood*
The teacher hit the boy at school yesterday with a stick

In (14), the subject sensei may be postposed to any of the positions marked with the arrow (↓) (but may *not* occur at x - before the object buik). Josephs' rule would predict that sensei would occur only at the end of the sentence, after kerrekar. In terms of this paper, (14) shows that Palauan is an OS language; the basic word order is VOSX or VOXS, where X is any oblique phrase.

Josephs also suggests a rule called Possessor-Preposing which accounts for some other cases of NP-initial clauses. Josephs makes no mention of the form of the verb in the preposing construction, as all of his examples in fact involve preposing from a clausal or phrasal subject, requiring the verb to have realis morphology. Unexplainable in Josephs' terms would be a sentence like that in (2c), where the initial NP is a possessor, yet the verb is irrealis. As for direct object preposing, no rule is suggested, as object-initial sentences are all analysed as passives (see the discussion of example (3)). We have seen that such sentences are actually transitive declaratives, and that passives have quite different properties (examples in (4)).

As we saw in (2), NPs of any grammatical relation may occur in clause-initial position. Following, essentially, the arguments in Waters 1979, it is easy to see how a single rule of topicalisation (from underlying VOS order) accounts for all the constructions termed by Josephs Subject-Shifting, Possessor-Preposing, and Passive, as well as other constructions illustrated here. The essential difference between possessor-preposed sentences and passives to Josephs is that the former have realis morphology and the latter irrealis; the analysis in section 3 shows that these facts should be analysed rather in terms of a rule that discriminates between subjects and nonsubjects.

A final observation from Palauan favouring the VOS hypothesis has to do with the fact that WH-words need not appear clause initially, or question initially, but may be found in situ - in whatever argument position they would normally occupy. We saw examples of this in (10) and in (13d). Some additional examples are found in (15). Note that in these examples the question word is a subject, and that it is *rightmost* in the clause.

(15)a. ngrirebet er a tebel a ngerang
R-3s-*fall* P *table* *what?*
What fell off the table?

b. ngngilmedii a ulaol a te'ang
R-3s-*clean* *floor* *who?*
Who cleaned up the floor?

c. akmedengei el kmo ngulu'ais a te'ang
R-1s-*know* *that* R-3s-*tell* *who?*
I know who told (the story)

That WH-words are not shifted *rightward* can be seen in sentences like (13d), where ngera *what?* is in the usual object position. In fact, question words may be found in situ in all NP positions, but only subject question words (or obliques) are found in the position illustrated in (15).

Looking at the whole range of Palauan sentences and questions, therefore, we find the evidence to be solidly on the side of the VOS hypothesis.

7. TYPOLOGICAL EVIDENCE

Having presented the language-internal evidence for a basic VOS word order, I would like briefly to turn to some typological considerations. As in any discussion based on typology, this section classifies Palauan with other verb-initial languages according to the existence of certain shared properties, properties that are central to the syntax. I will organise this part of my discussion around Keenan's (1978) article "The syntax of subject-final languages". Keenan describes a wide range of syntactic properties that characterise O-before-S languages.

7.1 Language family characteristics

Keenan observes that "subject-final languages normally occur in linguistic phyla in which verb-initial languages are common". Related to this is the generalisation that "subject-final languages are always verb-initial", to which Keenan notes certain exceptions. Verb-initial languages are the rule in the Malayo-Polynesian subgroup of Austronesian, to which Palauan belongs: the Philippine languages, the Polynesian languages, and languages like Malagasy, Chamorro, and Toba Batak are all verb initial.

7.2 Characteristics of the VOS type

Keenan points out that "SVO is a grammatical (although marked) word order in all VOS languages". We have seen above that this is true in Palauan. SVO order occurs not only in sentence elicitation when English (SVO) sentences are presented, but also in everyday conversation and in written stories. In addition, other NP-initial sentences occur, as in the examples in (2), (7), (8), and (9).

Another generalisation Keenan makes is that "if a language is subject-final then either transitive verbs ... agree with no full noun phrase ... or they agree with two noun phrases". Further, "if transitive verbs present agreement, then they have prefixal agreement with the subject and suffixal agreement with a nonsubject". All of the languages in Keenan's survey conform to these generalisations. Again these observations describe the case for Palauan verbs that have transitive forms. Verbs of perfective aspect agree with both subject and direct object; subject agreement is prefixed and object agreement is suffixed.

Keenan makes several other descriptive observations about the properties of subject-final languages. Although he unfortunately offers little discussion of the connection between word order and these properties, and although some of these properties may characterise languages with word order other than VOS, it is somewhat striking that practically *all* are characteristic of Palauan. The following are from Keenan, many are drawn from descriptions of Malayo-Polynesian languages, and all are true of Palauan.

In subject-final languages:

- there is relatively little nominal case marking
- prepositions rather than postpositions are found
- constituent questions can always be formed by putting the question word in preverbal position. In addition, "it is not necessary to front independent question words". Examples using either option have been given for Palauan.
- full NP possessors follow the head NP
- relative clauses present the head to the left of the relative
- there are no relative pronouns
- there is a passive form, which is morphologically marked
- there is no overt copula
- pronoun-retaining strategies in relative clauses are common, especially if the relativised position is constructed with a preposition.

The properties of Palauan syntax, in sum, correlate to a very high degree with the observed properties of other languages in its linguistic phylum, languages which are verb initial.[10]

8. DISCUSSION

In this paper I have concentrated on the syntax of Palauan sentences, and have offered no discussion of the pragmatics of the constructions in question. However, it can be noted that the obligatory retained pronouns and alternations in mood morphology described here constitute an efficient parser for locating the postverbal argument position which corresponds to a relativised NP or to a preposed topic or question phrase. If, on the other hand, the subject is understood as preceding the verb in the underlying order, this parser seems less motivated. This is not to deny that languages have redundant devices, but to point out that the machinery described here is well motivated given an underlying order in which both subject and object follow the verb, but not given an order like that of English.

In closing, I would like to comment briefly on the controversy surrounding the existence of O-before-S languages. It has been observed that languages with VOS order are relatively rare; by one count, less than 1% of the world's languages have OS order (Comrie 1981). Perhaps the most extreme claim is that of Parker (1980), who assumes that all human languages have a basic order in which the subject precedes the object. Pullum (1981) points out the weaknesses of the claims of Parker and other linguists as to the nonexistence or improbability of languages with certain rare word orders, and offers a list of 30 OS languages. The languages discussed in Keenan 1978 would add considerably to this list. Palauan is a language mentioned by neither, as it has been represented as being SVO. I hope not only that this paper has provided convincing arguments that Palauan is VOS, but that it will be taken as an illustration of the point, made

by Pullum and others, that classification of a language in terms of word order must be based on detailed language-specific analysis of a variety of clause types, and take typological considerations into account.

ACKNOWLEDGMENTS

I thank my Palauan consultants, Miriam Anastacio, Presley Etibek, Nina Ngirangol, Roy Ngirchechol, and Theodore Rengulbai, for their cooperation and interest. I thank also the staff of the Language Program and the Bilingual Education Project in Koror, Palau, especially Romana Anastacio, Isaias Ngirailemesang, and Mahensia Thomas, for their help in providing Palauan texts and English transcriptions. P. Kempis Mad, of the Belau National Museum, was an indispensable collaborator in all aspects of my research in Koror. This work has been supported by a Humanities Graduate Student Research Grant from UCSD, a National Science Foundation Doctoral Dissertation Research Grant, and a UCSD Academic Senate Grant to Professor Sandra Chung.

NOTES

1. I use the standard orthography approved by the Palauan Orthography Committee in 1972 and used in Josephs 1975; the only exception is the glottal stop, which I write ' where the orthography uses ch. It should be noted that this orthography is not in common use in Palau. The morpheme a marks major constituents and is not glossed, nor is tense glossed. Abbreviations are:

Dem	demonstrative	P	preposition
Im	imperfective	Pf	perfective
IR	irrealis	Ps	passive
L	linker	R	realis
p	plural	s	singular

2. See Waters 1979 for a number of arguments that the verb prefix in fact agrees with the *subject*, not necessarily with the preposed NP.

3. This is not the contradiction it seems: imperfective verbs in Palauan mark their objects with er, a preposition that also marks oblique objects, temporals, locatives, and possessors.

4. See Wilson 1972 and Flora 1974, as well as Josephs 1975 for more detail on Palauan morphology.

5. As noted above, the transitivity of an imperfective verb is marked by the infixed imperfective morpheme; transitivity of perfective verbs is marked by a suffix which agrees in person and number with the direct object.

6. In fact, agreement allows a gap in other positions, such as possessed NPs. For the sake of simplification I will concentrate on subjects and direct objects here. For greater detail see Georgopoulos 1985.

7. Oblique phrases, like the benefactive in (9), may either precede or follow the subject.

8. Waters justly points out, contra Josephs, that NPs other than subjects can be relativised. Waters further assumes that the relativised NP is usually topicalised within the relative clause, then deleted. These further assumptions appear unnecessary.

9. I refer informally to topic as a position of focus here, not implying any analogy to focus systems as found, for example, in Tagalog.

10. Discussion of a more concrete connection between these generalisations and underlying word order is, unfortunately, beyond the scope of this paper.

REFERENCES

CHOMSKY, Noam

1977 On *wh*-movement. In P.W. Culicover, T. Wasow and A. Akmajian, eds *Formal syntax*, 71-132. New York: Academic Press.

COMRIE, B.

1981 *Language universals and linguistic typology*. Oxford: Blackwell/ University of Chicago Press.

FLORA, Jo-Ann

1974 Palauan phonology and morphology. Ph.D. dissertation, University of California, San Diego.

GEORGOPOULOS, Carol

1985 Variables in Palauan syntax. *Natural Language and Linguistic Theory* 3/1:59-94.

JOSEPHS, Lewis S.

1975 *Palauan reference grammar*. Honolulu: University Press of Hawaii.

KEENAN, E.L.

1978 The syntax of subject-final languages. In W.P. Lehmann, ed. *Syntactic typology: studies in the phenomenology of language*, 267-327. Austin: University of Texas Press.

PARKER, Frank

1980 Typology and word order change. *Linguistics* 18:269-288.

PULLUM, G.K.

1981 Languages with object before subject: a comment and a catalogue. *Linguistics* 19:147-155.

ROSS, J.R.

1967 Constraints on variables in syntax. Ph.D. dissertation, MIT.

WATERS, R.

1979 Topicalization and passive in Palauan. MS.

WILSON, Helen I.

1972 *The phonology and syntax of Palauan verb affixes. Working Papers in Linguistics, University of Hawaii* 4/5.

Georgopoulos, C. "Palauan as a VOS language". In Geraghty, P., Carrington, L. and Wurm, S.A. editors, *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*.
C-93:187-198. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.187*

SURFACE ORDER IN THE STANDARD FIJIAN VERB PHRASE

David G. Arms

0. INTRODUCTION

This paper attempts to describe some of the intricacies of surface order within the Standard Fijian verb phrase, and also how these order phenomena relate to the larger sentence. Although a little treatment will be given to underlying structure, the paper will largely be restricted to providing an overview of what actually occurs on the surface. A fuller accounting for surface occurrences in terms of underlying relationships must await another occasion and a more extensive grammatical study.[1]

1. BASIC ORDER WITHIN THE VERB PHRASE

The core of the verb phrase in Fijian consists of a subject pronoun, verb and optional object[2] - this object being a pronoun, proper noun or phrase, or in the indefinite construction (Hazlewood 1850:31; Churchward 1941:19; Arms 1974: 60)[3] a common noun or phrase,

(1) au raici ira
I see them

However, a few elements may occur before the subject pronoun, a large set of elements may occur between the subject pronoun and the verb, and an even larger set may occur after the object. Some of these can be termed adverbs. Following terminology being introduced by Schütz (1985:chapter 7) I will term the remainder "(verb phrase) markers".

1.1 On the following two pages is a list of the various elements that can occur in the verb phrase - though not all at once, of course! The significance of the four columns and the meanings of the symbols employed are explained at the end of the list.

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 199-230. *Pacific Linguistics*, C-93, 1986.

Slot	Element	Meaning	Modifies
1.	Conj.1	Conjunctions: ia *but*, ka *and*, se *or*	-
2.	Conj.2	Conjunctions: ni *because, when, that* (realis),	-
		me *that* (irrealis), kē *if, then*, dē *lest, in*	-
		case, perhaps	
3.	Subj.P.	Subject Pronouns: au *I*, o *you*, etc.	-
4.	ā	past tense	C
5.	sā	*now, as of now*	C
	se	*as yet, still, meantime, already*	C
6.	na	future tense	C
7.	rairai*	*seem to*	C
8.	qai	*then* (time sequence)	C
9.	mani	*then, so, in the circumstances, finally*	C
10.	tei	*first*	C
11.	baci	*yet again, yet*	V
12.	bau	*somewhat, rather, at all*	M,V
13.	rui	*too, extremely*	M,V
14.	dau*	*customarily, habitually, frequently*	V
15.	dui	*respectively, individually*	P
16.	mai*	*come and*	V
	lai*, laki*	*go and*	V
17.	via	*want to, practically, almost*	V
	viavia	*practically, almost*	V
	vakarau*	*about to, ready to*	V
18.	tekivū*, vakatekivū*[4]	*begin to*	V
	tini*	*end up*	V
19.	vei-	reciprocal, etc.	V
20.	vaka-	causative, etc.	V
21.	O.Pref.	Other Prefixes: nā-, tā-; ca-, ka-, etc.[5]	V
22.	Main V.	Main Verb	-
23.	Tr.E.	Transitive Endings: -ci, -gi, -ki, etc.	-
24.	Object	Object Pronouns: au *me*, iko *you*, etc.	-
		Other Noun or Noun Phrase	-
25.	Mod.V.	Modifying Verb	V
26.	Adverb	Adverb	V
27.	lesu*	*back*	V
28.	lō	*secretly*	V
29.	yādudua, yārurua, ...	*individually, in twos, ...*	P
	yādua*, yārua*, ...[6]	*each one, each two, ...*	P
30.	vata	*together*	P
31.	kaya	*with him/her/it*	-
32.	tani*	*away*	V
	laivi*	*away*	V
33.	cake*	*up(wards), above*	M,V
	sobu*	*down(wards), below*	M,V
34.	koso*	*prematurely, on the way, in the course of things*	V
35.	oti*	*finished, already*	V
36.	rawa*	*get to, manage to, achieve, be able to*	V
37.	makawa*	*a long time ago*	V
38.	taumada	*first, foremost*	P,V,C
39.	duadua*, ruarua, ... kece	*alone, both, ... all*	P
	taudua*, taurua*, ... taucoko*	*just one, just two, ... the whole, all*	P

Slot	Element	Meaning	Modifies
40.	soti, sō*[7]	*so much*, pluraliser	P
41.	wale*	*only, merely, simply, fruitless(ly)*	M,V
42.	dina*	*truly*	M,V
43.	saka*	*sir/madam*	C
44.	sara	*immediately, very, really, precisely*	M,V
45.	tale*	*again*	V
	tiko*	*stay*,[8] currently-operative action/state	V
	toka*	*squat*, short-enduring action/state	M,V
	tū*	*stand*, long-enduring action/state	V
	koto*	*lie*, long-enduring action/state	V
	nō*	*lie*, long-enduring action/state	V
	voli*	*around*, currently-operative action/state	V
47.	bagi	*indeed!, after all*	P,V,C
48.	mada	*mind, first, if you please*	P,V,C
49.	beka	*perhaps, may be*	P,V,C
50.	gona*	*therefore, hence*	P,V,C
51.	lī	*pray!, indeed!*	P,V,C
52.	rī	*surely!, indeed!*	P,V,C
53.	gā, lā[9]	*just, anyway, only*	P,M,V,C
54.	mai*	*in this direction, hither, and come*	V
	yani	*in that*(2) *direction, thither*(2), *and go*(2)	V
55.	kina	*at/to/by/with it/them*	-
56.	oqō*, oqoka*	*at this time/place, this*	C
	oqori*	*at that*(2) *time/place, that*(2)	C
	oyā*, gona*[10]	*at that*(3) *time/place, that*(3)	C
57.	vuā*	*to him/her/it*	-
58.	Adverb	Adverb	V

1.2 Column one of the above list numbers the various slots for ease of reference. Items in the same slot are mutually exclusive. Other co-occurrence restrictions are indicated in 3.ff. The slots have been placed in sequence according to the surface order that applies when a number of the above elements co-occur. In some cases alternative orderings are possible. These are treated in 2.ff.

1.3 Column two lists the various potential elements of the verb phrase. Milner (1972:94 and 116) was the first linguist to study in detail the ordering of the verbal markers.[11] There are quite a number, however, that he did not include in his ordered lists, although some of these he did treat elsewhere in his grammar. As far as I am aware, the list I have provided above in 1.1 contains the full set of verbal markers.[12] The only elements I am not regarding here as markers are the verbs, the objects and the adverbs.[13] Asterisked markers are forms which may occur with related meaning in a capacity other than as markers.

1.4 Column three provides glosses for the various elements. These glosses do not claim to be semantically exact and it is regrettably not possible within the confines of this presentation to develop the semantics further. However, more insight into the meanings of the various markers can be gleaned from the examples provided throughout the paper. All slots are represented somewhere in the paper.

1.5 Column four suggests what each marker modifies. I use the term "suggest" designedly as a lot more work is necessary in this area (cf. also 6.ff). The

entries in column four indicate that the element in question would seem to modify: the entire clause (C), the verb (V), a pronoun (P), a marker or an adverb (M). Some markers have multiple entries, indicating they can modify different things on different occasions.

2. ORDER VARIATIONS IN THE VERB PHRASE

When a number of elements from the list of 1.1 co-occur, they *must*, in most cases, do so in the order presented; in other cases, the order presented is only the most *usual* one, or at least a very common one. The possible alternative orderings I am aware of are presented below. They do not change the meaning except where noted.

2.1 The subject pronoun may invert around the first occurring marker if this be ā, sā, se, qai or mani (slots 4, 5, 8 and 9 on the list).[14] Where ā sā, ā se, sā na, or se na occur however, the first person exclusive pronouns (and only these) invert around both elements if inversion takes place.[15] For other combinations (e.g. se qai, sā mani), the first person exclusive pronouns may invert around one marker only or both. All these pronoun inversions are a stylistic property of more lively discourse. Inversions involving sā and se are the most common.

2.2 rairai (no.7) may also occur after qai or mani (nos.8 and 9). It may even occur after baci or bau (nos.11 and 12) but this is not common. The use of rairai as a marker would clearly seem to derive from one of its uses as a verb.

(2)a. e rairai ni ra vinakata tale beka gā me ...

b. e rairai era vinakata tale beka gā me ...
it seem that they want-it again[16] *perhaps just that*

c. era rairai vinakata tale beka gā me ...
they seem want-it again perhaps just that ...
It seems (that) they may also want to ...

The three sentences are synonymous and seem to represent three progressive stages of development.

2.3 mani (no.9) may precede qai (no.8). Either ordering is reasonably common. These two markers are not mutually exclusive as is implied in Milner's listing.

2.4 rui is the most versatile of all the preverbal markers. From its common position (no.13) it may move ahead of any of the preceding markers up to and including qai (no.8). It may also follow any of the markers of slots 14 to 16.

2.5 dau (no.14) is also quite versatile. It too may occur ahead of the preceding markers up to and including qai (no.8). These alternatives are not frequently encountered however, except for bau and dau for which either ordering, bau dau or dau bau, is common.

2.6 dui (no.15) may invert around slot 16 (mai, lai and laki) with the possibility of a slight meaning difference.

(3)a. era dui mai taura na nodra ivola
they respectively come-and take-it c-art.[17] *their book*
They respectively came and took their books.

b. era mai dui taura na nodra ivola
they come-and respectively take-it c-art. *their book*
They came and respectively took their books.

In example (3a) the individuation of the action covers the coming as well as the taking of the books. In example (3b) this is not stated: they may have come as a group and then individually taken the books.

More rarely dui may invert around slot 17 without change of meaning.

2.7 mai and lai (and laki) - that is slot 16 - may also invert around slot 17. Indeed when via means *want to*, via mai, via lai and via laki are probably the more frequently encountered orderings. But both orderings are common and are often associated with a meaning difference.

(4)a. na nona via lai sarava
c-art. *his want-to go-and watch-it*
his wanting to go and watch it

b. na nona lai via sarava
c-art. *his go-and want-to watch-it*
his proceeding to want to watch it

The meaning of lai is more literal in (4a), figurative in (4b).

2.8 Although only one "main verb" slot has been entered on the list of 1.1 (slot 22), this slot may in fact be filled by two verbs which either form a compound verb or are simply two verbs in conjunction.

In the case of a compound verb the transitive ending, if any, will occur on the second verb. If the object of the compound verb is third person unmarked however, this object may occur (along with the transitive ending of course) on the first verb as well - that is, slots 23 and 24 may in this case be appended to both verbs.

(5) eda qai mai via rai(ca) lesuva
we-incl. *then come-and want-to see(-it) go-back-to-it*
We next wanted to look back over it.

When the verbs in this slot are simply conjoined, the transitive ending and object slots will also be repeated if both verbs are transitive.[18] In written texts, ka *and* is inserted as well.[19]

(6) era ā saravi keirau (ka) vakarogoci keirau tiko
they past *look-at we*-2-excl. *(and) listen we*-2-excl. *stay*
They were looking at and listening to the two of us.

2.9 The transitive ending -taki of slot 23 occasionally occurs along with another transitive ending if this latter is employed in the reciprocal or dispersive form of a verb.

(7) e dodonu me Ø[20] veilewaitaki o koya
it right that he try p-art. *he*
He should be brought to trial.

(8) Ø sā vakaveisēyakitaki ira na viavialevu
he now scatter them c-art. *proud*
He scatters the proud.

Such uses, however, would appear to be relexifications and it is not therefore necessary to regard them as true exceptions to the normal order pattern.

2.10 One idiomatic use of -taka (slot 23) allows it to be disjoined from its verb and occur later in the overall order. This is the use of -taka, usually with na cava *what*, in the sense of *for, reason.*

(9) na cava era ā lako kece taka mai
c-art. *what they* past *go all for hither*
For what reason did they all come?

It is not feasible to regard lako kece here as a special lexification. Rather, taka would appear to be functioning like an additional marker. I have not given it a separate slot however in the overall order, as it is in general mutually exclusive with the transitive endings and objects. However, it is not mutually exclusive with the indefinite construction "object" nor with the transitive endings when these are employed in the reciprocal or dispersive forms of the verb.

(10) na cava o gunu yaqona taka tale tiko
c-art. *what you drink kava for again stay*
Why are you drinking kava again?

(11) na cava era veiraici taka toka vokadedē
c-art. *what they* recip.-*look for squat long-time*
Why are they looking at each other for such a long time?

It seems that for some speakers, this use of taka can occur well out in the order, even perhaps after dina (slot 42). For others, these unusual orders (such as example (9), but not (10) or (11)) are questionable.

2.11 Like the verb slot, the object slot (no.24) may have a double filler if this is a proper object (more particularly a locative one) or an "indefinite object".

(12)a. e na sikovi Viti kei Rotuma bagi
he future *visit Fiji and Rotuma indeed*

b. e na sikovi Viti bagi kei Rotuma
He will indeed visit Fiji and Rotuma.

(13)a. e sā gunu bia kei na waini tiko
he now drink beer and c-art. *wine stay*

b. e sā gunu bia tiko kei na waini
He's drinking beer and wine.

Examples (12b) and (13b) are alternative orderings where the conjoined phrase is moved outside the verb phrase altogether. This is the preferred ordering for example (13), but both orderings are fully acceptable for examples like (12).[21]

2.12 Slot 25 represents a very common modifying device whereby one verb is placed after another, with or without intervening object, in a kind of adverbial function.

(14) e sauma lagalaga lesu o koya
he answer-it sing back p-art. *him*
He answered in a singing tone.

Sometimes it can be difficult to be sure whether one is dealing with a construction of this type or of the compound verb type (cf. 2.8).

(15) ka sā davo wāwā koto mai kina o Qeila
and now lie wait lie hither at-it p-art. *Gael*
Where Gael lay waiting.

However, it should hardly surprise us that there would be some overlapping of construction types. In any event, the adverbial-like construction of slot 25 is very common, and it is also quite common for reduplicated forms to fill this slot.

2.13 By the term "adverb" in slot 26 I am referring to vaka- forms such as vakatotolo *quickly*, vakavinaka *well*. There is another adverb slot (no.58) at the very end of the verb phrase. These two slots are normally to be considered as alternative orderings, but occasionally both are filled at the same time if the two adverbs concerned are serving disparate functions.

(16) e kauti koya vakavōleka sara tale vakasaurī vuā na wekana
it carry her closely very again suddenly to-him c-art. *her-relative*
It suddenly brought her very close again to her relative.

More than one adverb can fill either individual slot at the one time.

(17) e qai rabocaka vakabābā vākaukaua sara gā e mata i Jone
he then swing-it sideways hard very just in face of John
He then swung it real hard across John's face.

If the two adverbs are serving similar functions, they may, in the written language particularly, be conjoined by ka *and*.

(18) nī wilika mada vakamatata ka vakadodonu
you read-it please clearly and correctly
Please read it clearly and correctly.

On rare occasions the adverb may move outside any of the slots up to and including no.34 (koso) yet not be filling slot 58.

(19) e ā lako tani vakatotolo tale gā o Pita
he past *go away quickly again just* p-art. *Peter*
Peter also went away quickly.

But in general such sentences are stylistically clumsy or unacceptable.

The motivation for choosing between slots 26 and 58 is very often stylistic or the need for clarity. In addition, choice of slot 26 may allow the adverb to be itself further modified more easily (cf. 4.15ff).

Some adverbs when they occur in slot 26 (or its variant above) - but never in slot 58 - may occur without the vaka- with no change of meaning.

(20) e cakava (vaka)vinaka tiko o koya
he do-it well stay p-art. *him*
He is doing it well.

Since the adverbial element is in this case reduced to its basic root, one could argue that such occurrences are really manifestations of slot 25. Certainly if a non-vaka- and vaka- adverb occur together, the former precedes. Nevertheless, I prefer to regard instances such as vinaka in example (20) as filling slot 26 in view of their relationship with the vaka- forms. However, we could have here another example of overlapping constructions.

2.14 lō (no.28) and the numerals of the following slot (no.29) may occur, albeit rarely, outside any of the markers up to and including koso (no.34).

2.15 vata (no.30) often occurs in construction with kei *and, with* or kaya *with him/her/it.*[22] When it does so, it may occur totally outside the verb phrase with these elements. However, it may also occur inside the verb phrase. In this case kei, if it occurs, must remain outside the verb phrase whereas kaya, if it occurs, must go with vata inside the verb phrase. Indeed when vata and kaya both occur, nothing may ever come between them. When vata occurs without kei or kaya, and conversely when kaya occurs without vata, they must occur inside the verb phrase. The possibilities therefore are as follows (using tiko as an example of another marker in the verb phrase):

(21) a. era sā lako tiko vata kei koya
they now go stay together with him

b. era sā lako tiko vata kaya

c. era sā lako vata tiko kei koya

d. era sā lako vata kaya tiko
They are going together with him.

e. era sā lako vata tiko
They are going together.

f. era sā lako kaya tiko
They are going with him.

Occasionally vata and/or kaya may, like slots 28 and 29, invert around the subsequent particles up to and including no.34 (koso).

2.16 The markers of slots 32 to 34 may occasionally change order with each other, but cf. 3.13. For some speakers at least, koso may occasionally occur after oti or rawa (nos.35 and 36).

2.17 rawa (no.36) occasionally occurs after slots 37 to 39. rawa occurs as a main verb as well as a marker. Unlike the case of rairai though (cf. 2.2), this order variation is associated with a semantic difference. Just as there is

an important distinction to be made between a clause introduced by rawa ni or rawa me (cf. Milner 1972:62), so there is a distinction between both of these uses and rawa used as a marker. Whereas the former indicate a mere possibility or a permission (or wish) respectively, the latter indicates that something is achieved or to be achieved.

(22)a. e a rawa ni raica na vuniwai
he/it past *able that see-him* c-art. *doctor*
It was possible for him to see the doctor.

b. e ā raica rawa na vuniwai
he past *see-it able* c-art. *doctor*
He got to see the doctor. OR *He was able to see the doctor.*

The difference in meaning here is important. You can add the clause ia, e sega ni raici koya *but he did not see him* to example (22a) but not to (22b), as that would contradict the very statement just made. The translation *be able* for rawa must be handled with care due to its ambiguous use in English: *be able = be possible, can; be able = achieve, effect.* Note that imperatives are possible with rawa as a marker.

(23) Ø katalau rawa daru qai lako
you breakfast able we-2-incl. then go
Have your breakfast and then let's go.

2.18 makawa (no.37) and taumada (no.38) may possibly occur also in the reverse order, and either of them may occur before rawa (no.36). In addition, taumada may sometimes invert around slots 39 and 40. This is probably done to mark contrast.

(24) au na raici ira kece taumada; oti, au na qai raici ira
I future *see them all first finish I* future *then see them*
yādudua
individually
I will see them all first; after, I will see them individually.

2.19 wale (no.41) and dina (no.42) can occasionally occur before kece but not, as far as I am aware, before other members of slot 39.

2.20 saka (no.43) is a marker used to show respect. When very great respect is being shown it may occur a number of times in the same sentence in different phrases or even outside of (after) such phrases. In this latter case or when saka is at the end of a phrase, it always receives a distinctive lowered intonation. When followed by some other element of the same phrase however, saka usually loses this distinctive intonation and is absorbed into the normal phrase intonation pattern.

2.21 gona (no.50) can occasionally be found to occur after lī and rī (nos.51 and 52) still bearing the same meaning.

2.22 Slots 56 and 57 sometimes occur in the reverse order. It may be questioned why these elements are regarded as part of the verb phrase at all. It is true that they do also occur outside the verb phrase. Indeed, if oqō, etc. is expanded into an equivalent phrase (e.g. e na gauna oqō *at this time*) or if vuā is further

specified (e.g. vuā na tamana *to his father*) they must occur outside it. However, when they occur unembellished there is a strong tendency, though by no means a rule, for them to move inside an adverb of slot 58 and to be treated phonologically as part of the verb phrase.

(25) au se qai cakava oti wale ga oqo vua vakavica
I as-yet then do-it finish only just this to-him several-times
I've already just done it to him a number of times.

2.23 Comment

The number of order variations given in this section looks forbidding. But when we consider the enormous number of markers involved, it would be surprising if there was not some variety of usage. Indeed what is remarkable is the degree of regularity and strictness that exists in the ordering of the markers. Another remarkable thing, of course, is the extraordinary richness of expression and nuance that can be conveyed in the Fijian verb phrase. The range of possible options is very full as there are few enough co-occurrence restrictions as we will see below (cf. 3.ff).

An interesting additional way of looking at the list of 1.1 is to consider it in terms of zones. The markers tend to group semantically (though there are exceptions) and according to what other markers they may change order with.

One possible zone division would be the following:

(i) Slots 1 and 2 are the verb phrase introducers (conjunctions) and do not allow order variation.

(ii) Slots 3 to 6 are the early markers which cannot be preceded by later markers but which do allow some inversion of the pronouns (slot 3). Apart from the pronoun slot itself, the semantics of these markers is tense/aspect.

(iii) Slots 7 to 9 are an inbetween set. They share the possibility of some form of pronoun inversion but also allow order variation with some of the following slots. Their semantics too have relationships both ways.

(iv) Slots 10 to 15 are the middle set which admit a considerable amount of order variation but also some mutual exclusiveness.

(v) Slots 16 to 18 are becoming very verb-like in their appearance and semantics.

(vi) Slots 19 to 21 are the prefixed elements that enter into a particularly close relationship with the verb.

(vii) Slots 22 to 24 are the verb and object elements.

(viii) Slots 25 to 34 (plus slot 58) contain elements that are very adverbial in nature. There is a fair range of order variation possible among them, although in fact some occur together only rarely.

(ix) Slots 35 to 42 are a set which does allow a little order variation. Semantically they all give some indication of fullness or completeness.

(x) Slots 43 to 55 are a very important set that allows no order variation apart from that mentioned in 2.21 for gona. Many of these are of very common occurrence.

(xi) Slots 56 and 57 are a small set that can vary order and move outside the verb phrase altogether.

Other zone divisions can also be made. Thus slots 47 to 52 are an interesting set that seem to emphasise or mollify certain grammatical moods, respectively: bagi exclamatory, mada soft imperative, beka dubitative, gona explicative, lī emphatic, rī confirmative.

A fuller study of the properties of such zones would yield interesting results. In some zones there are some relative co-occurrence restrictions, but few that are absolute (cf. 3.ff). The ordered listing of 1.1 is an extremely important part of Fijian grammar. Even granted the variations described in this section and others I may have missed, an appreciation of the ordering requirements presented here is basic to proper use of the markers and to understanding their structural relationships.

3. CO-OCCURRENCE RESTRICTIONS AMONG VERB PHRASE MARKERS

It is not possible here to consider all the possible co-occurrence restrictions that may exist between the multitude of markers and the fillers of the verb slot. For one thing this would presuppose a better understanding of Fijian word classes than currently exists (although the Fijian Dictionary Project is doing a lot to remedy this). For another, the number of markers makes the task unmanageable here. Though I will allude to some such co-occurrence restrictions in passing, my main concern will be co-occurrence restrictions among the markers themselves.

3.1 One type of co-occurrence restriction is that induced by length. This in turn is governed no doubt by what constitutes a phonologically manageable phrase, by what makes practical sense and is not semantically over-packed, and by what is stylistically desirable. Obviously, with such a huge number of markers to choose from, the speaker will select only a very limited set at any one time. The most I have observed occurring before the verb is six.

(26) ni ra ā sā qai dui lesu tale i vale
for they past *now then respectively go-back again to house*
for they had then each returned home

One does not often encounter this many markers before the verb, but four or five are common.

The situation is similar after the verb. One would only rarely find six markers in this position (example (27)) but as many as four or five are common. There are some favourite combinations (cf. 4.ff and 5.ff) which tend to make the build-up to this number more easy.

(27) ni sā tagi dina sara tū mada gā mai o Nina
for now dry truly very stand mind just hither p-art. *Nina*
for Nina was really just crying her eyes out to us

The fact that these numbers of markers can occur before or after the verb does not mean that you can have such numbers both before and after the verb at the same time. This would be just too much phonologically, semantically, and stylistically. Example (27) above has a total of eight markers in all. The highest number I have observed is given in example (28) below. The total number of markers is ten if oqō is included in the count.

(28) ni keirau ā se qai sota kaya oti koto gā oqō
for we-2-excl. past as-yet then meet with-it finish lie just this
for we had only just met with it then

3.2 Other co-occurrence restrictions derive chiefly from the fact that particular markers are covering the same or a closely related semantic field and distinct choices must be made. Many of these co-occurrence restrictions are already implicit in the list of 1.1. Where two or more markers are entered in the same slot they are mutually exclusive. Other co-occurrence restrictions, some absolute some relative, are given below.

3.3 The conjunction ia *but* of slot 1 only dubiously belongs there since it is not really part of the verb phrase phonologically. It has been included in order to complete the inventory of conjunctions. It differs from ka *and* and se *or* too, in that it is purely coordinative whereas these latter also function as subordinating conjunctions, ka to introduce relative clauses[23] (also in the ka ni construction, cf. Milner 1972:76) and se to introduce indirect questions (... whether ... or ...).

3.4 The conjunctions se (slot 1) and ni (slot 2) rarely occur together, and certainly never when the se is conjoining what look like two ni clauses.

A rather important and common construction that does not seem to have been written about previously is the following:

(29) au sega ni bau kilai koya vakadua se me u rogoca mada e liu
I not that at-all know him once or that I hear-it mind at before

na yacana
c-art. *name-his*
I hadn't met him at all ever or even so much as heard his name before.

There are two clauses dependent on the negative sega, but the second is cast into a different mood by the introduction of me. If on the other hand the conjoining was straightforward, only se would be used, me and mada dropping out; ni would not replace me.

3.5 It will be noticed that I have not included kēvakā in the conjunction slots. This is because I consider it as better analysed as two words, kē and vakā *be like it* (vakā also occurs separately in me vakā *like, for example*). True, nothing may come between kē and vakā, but the occurrence of markers after the vakā (e.g. kē vakā beka *if perhaps*) would seem to recommend consideration of vakā as a verb in this combination.[24]

An important construction with kē vakā that has been omitted in the grammars is the fact that kē vakā may introduce a clause beginning with me, not just an ordinary indicative one.

(30) a. kē Ø vakā e lako mai
if he like-it he go hither
if he comes

b. kē Ø vakā me Ø lako mai
if it like-it that he go hither
if he should come

The translation captures well the difference in nuance obtained by using me. A greater sense of uncertainty is involved.

When the event concerned is completed or is currently going on, the use of me clearly indicates that the condition is contrary to fact (example (31b) below). The indicative clause (31a) on the other hand may be interpreted as either an open or a contrary to fact condition. Even where the indicative clause is contrary to fact, however, the use of me conveys a greater degree of remoteness or uncertainty comparable to example (30b).

(31)a. kē Ø vakā e ā tiko e na bose kē Ø ā vinaka
if it like-it he past *stay at* c-art. *meeting then it* past *good*
If he was at the meeting, it was good. OR *If he had been at the meeting, it would have been good.*

b. kē Ø vakā me Ø ā tiko e na bose kē Ø ā vinaka
if it like-it that he past *stay at* c-art. *meeting then it* past *good*
If he had been at the meeting, it would have been good.

me is used in this way after kē vakā but not, it would seem, after kē alone. This is further support for analysing the vakā in kē vakā as a verb. If kē vakā were analysed as an alternative conjunction to kē, it would then violate the co-occurrence restriction between members of slot 2. Note, however, the following:

(32) kē u ā vesuki koya mai se me u ā vanai koya mate gā
if I past *catch him hither or that I* past *shoot him dead just*
if I had caught him or even shot him dead

Although the first clause is introduced by kē and the indicative, the second has a mood change, being introduced by me.[25]

3.6 ā and na, the two tense markers (slots 4 and 6), as one might expect, cannot co-occur.

3.7 tei (no.10) cannot co-occur with rairai (no.7). I have not encountered, nor been able to elicit, convincing examples of it with rui (no.13) or dau (no.14) either. Occurrence with baci (no.11) and bau (no.12) are not common. tei does not occur either in nominalisations (cf. 5.1). These severe restrictions on tei no doubt stem from its imperative nature. It is used in sentences expressing command, intention or volition and indicates that the action should 'now forthwith' commence, often with the implication that it is to precede something else or at least that it is being done in the awareness that some other action or state will follow. I have left tei in position no.13 as informants who accept the rare combinations prefer or require the order given. Milner too (1972:116 note) explicitly considered some of the combinations.

3.8 baci (no.11) and bau (no.12) rarely, if ever, co-occur. It can be questioned then why they occupy separate slots. However, there are some marginal (elicited) sentences where this order seems preferred. Milner too (1972:116 note), as for tei above, evidently found some support for this order.

3.9 rui (no.13) can only modify an element that expresses degree. If the verb itself does not express degree, then rui can only occur in a sentence containing that verb if some adverb or marker expresses degree.

(33) e rui cici vakatotolo
he too run quickly
He ran extremely fast.

The sentence *e rui cici alone would be ungrammatical.

3.10 dui (no.15) requires a non-singular subject in order to occur. Remember, however, that e is an *unmarked* third person pronoun, not necessarily singular.

(34) na kā e dui baleti ira na gone
c-art. *thing it individually concern them* c-art. *child*
the things that individually concern the children

3.11 via (no.17) in the sense of *want to* cannot co-occur with oti (no.35).

3.12 The first two prefixes listed in slot 21, namely nā- and tā-, are the only ones that can occur with the prefixes of the preceding slots, vei- and vaka-; and indeed if they are to occur with vei-, vaka- must intervene. Another prefix sau- can actually occur ahead of tā- but I have not provided it with a separate slot because of its restricted use.[26]

Other prefixes of slot 21 would be the passive prefix lau- and the spontaneous prefixes ca-, ka-, etc. (Arms 1974:54 and 72); also the prefixes kolai *almost*, tabu *not* and tawa *not* which occur only with an extremely limited number of roots. Finally, there is the prefix (tau)yā- *each* which occurs only with numerals (including vica *how many*), and the form lewe (or lē) *person* which is not usually regarded as a prefix but as a separate word and which occurs with quantifiers only. All of these prefixes are mutually exclusive with vei- and vaka- (slots 19 and 20).

3.13 The co-occurrence of any of the fillers of slots 32 to 34 is very rare.

3.14 taumada and mada (nos.38 and 48) do not co-occur, probably because of their shared phonological and semantic properties.

3.15 bagi and beka (nos.47 and 49) rarely, if ever, co-occur.

3.16 mada and beka (nos.48 and 49) are registered by Milner (1972:94) as not co-occurring. By and large this is correct, but there do seem to be rare occasions when the combination is possible for some speakers.

(35) me Ø tei dua mada beka na memu bilo tī mo qai mākutu
that it first one mind perhaps c-art. *your cup tea that-you then sharp*
First have a cup of tea perhaps, so as to be sharp.

3.17 As already stated in 2.13, the two adverb slots (nos.26 and 58) are rarely filled at the same time.

3.18 Comment

The above co-occurrence restrictions among verb phrase markers are all I can provide at this time. Undoubtedly further research will uncover other such restrictions, and of course I have barely even touched the question of co-occurrence restrictions between the markers and the verb itself. Nevertheless, one

does get the impression that there is a high degree of freedom and compatibility among the markers, and between the various markers and the verb.

One problem in this area is the interference of dialect. As a standard language, Standard Fijian is spoken by a host of speakers who also know at least one other dialect. Undoubtedly dialect features are often carried over into the standard language, and at the margins of grammaticality it can be difficult to determine what is standard and what isn't. The problem is further aggravated by the fact that speakers at times spontaneously declare certain forms to be incompatible whereas in reality it is just that they have been unable to envisage the semantic scenario which would allow their co-occurrence.

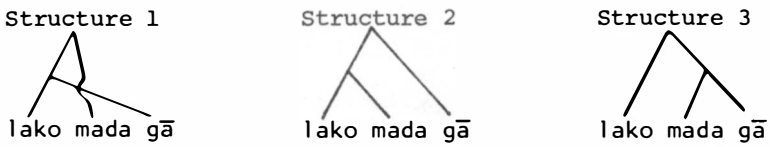
Finally it should be noted that in this section I have been checking the compatibility of one marker with a second. I have not checked on the compatibility of any three or four markers together. It could be expected that co-occurrence restrictions at this higher order would in general derive from the incompatibility from two members of the combination, but other factors such as semantic over-packing may play a part.

4. CO-OCCURRENCE TENDENCIES AMONG VERB PHRASE MARKERS

In column four of the list of verb phrase elements in 1.1, I have made some suggestions as to what I believe the various verb phrase markers habitually modify. This is basically an impressionistic judgement provided to give a broad idea of the likely structural relationships within the sentence. As indicated in 1.5, this categorisation should not be given too much weight.

Just how the markers modify the clause or verb cannot be gone into here. Some markers, however, occur rather frequently together and it can be suspected they enter into a specially close relationship with one another. It is these markers that will be the focus of attention in this section.

Three ways in which the markers could enter into relationship with each other (and with the verb, etc.) would be as illustrated below (where Ø lako mada gā = *go anyway*).



In structures 1 and 2, one marker modifies the verb first, then the other modifies the joint combination. In structure 3, one marker modifies the other first and then the combination modifies the verb.

It has been traditional to assign separate meanings to some of the combinations of markers. This may be a good device to assist the language learner, but linguistically the semantics of most of the combinations are adequately accounted for in terms of the constituent structures illustrated above. The combinations are not as a rule to be regarded as idiomatic,[27] except perhaps in the sense that they realise one particular constituent structure rather than another.

4.1 tei ... mada (nos.10 and 48). When tei *first* occurs, mada *please* very frequently, but by no means always, occurs with it. This is probably because

tei is a polite word, and it is very common in polite contexts to include mada as a softener when expressing commands, wishes, or intentions.

4.2 baci ... tale (nos.11 and 45). Similarly, when baci occurs, tale frequently, but not always, occurs with it. The combination of the two makes the iteration stronger: *yet again.*

4.3 rui (no.13) can directly modify an adverb or another marker expressing degree (thus realising structure 3) as well as being able to modify directly certain verbs (cf. 3.9).

4.4 mai ... yani (nos.16 and 54). When the preverbal mai occurs along with the postverbal yani, a meaning like *proceeded to* is often conveyed.

(36) e mai vākuvukuvu toka yani
he come-and smoke squat thither
He proceeded to smoke.

This sentence was used after describing all the preparations that went into making the cigarette. The usage would appear to be a figurative way of employing the semantics of the markers. The action 'comes' from one direction (mai) and carries 'yonder' (yani). The fact that these two markers occur in a sentence does not mean they have to be interpreted in this figurative way. Their more literal meaning may apply.

4.5 When the fillers of slots 17 and 18 (via, etc.) occur with those of slot 46 (tiko, etc.), the former enter into construction with the verb first and the combination is modified by tiko, etc. (that is, structure 2 applies), but it can be proposed that in at least some cases the two slots enter into construction first and only then modify the verb (that is, structure 3 applies). Compare the two synonymous sentences below.

(37)a. era sā tekivū vosavosa tiko mai
they now begin talk stay hither

b. era sā tekivū tiko me ra vosavosa mai
they now begin stay that they talk hither
They are beginning to talk.

4.6 vata kaya (nos.30 and 31). The special relationship between these two markers has already been discussed in 2.15.

4.7 When oti (no.35) occurs, sā very commonly, but certainly not necessarily, occurs with it. This correlation derives from the internal semantics of the markers in question: sā often indicates a new state of affairs that has come about *now*, and oti that a certain action is *finished.* However, when oti occurs in a clause introduced by ni *when*, the sā is frequently omitted. The time correlation is sufficiently borne by the conjunction ni itself (cf. Arms 1978:1268).

4.8 se ... qai ... oti ... gā (nos.5, 8, 35 and 53). When any one or any combination of the first three markers here occurs with gā, the combination means *just* (referring to time), there being different nuances according to which particular markers are employed.

(38) e se qai wilika gā e rua na matanivola
he as-yet then read-it just it two c-art. *letter*
He had just read two letters (of the alphabet).

The combinations se qai and se qai ... oti alone (without gā) can also have much the same meaning.[28]

4.9 The above combination of gā with another marker or markers is only one of several such combinations that gā enters into. Starting with the markers closest to gā and working back towards the verb, the combinations traditionally given are:

mada gā	*mind, even*
tiko gā, etc.[29]	*still, keep on*
tale gā	*also, too*
sara gā	*absolutely, exactly*
dina gā	*really and truly*
wale gā	*only, just*
kece gā	*really all*

These combinations will be treated in the following paragraphs.

4.10 When any one of the markers kece,[30] wale, sara, or tale occur in the same phrase as gā, they enter into construction directly with it - structure 3. Note, however, that a somewhat different rule applies when more than one of these four markers occur together, cf. 4.14.

4.11 dina gā is not in fact a commonly found combination. When these two markers do co-occur it seems they are free to function as a unit (structure 3, with the meaning given above) or as independent markers (structure 2).

4.12 The combinations tiko gā, toka gā, etc. (slot 46 plus gā) pose a special problem. They often provide the reading given in 4.9, namely *still, keep on.* However, tiko, toka, etc. also occur as main verbs and, when gā is added, can have much the same meaning.

(39)a. nī gunu tiko gā
you-pl. *drink stay just*
Keep on drinking.

b. nī tiko gā
Stay on. OR *Stay anyway.*

It is tempting to propose, as traditional treatments imply, that structure 3 applies to example (39a); that is, tiko and gā enter into construction and then jointly modify gunu. But it is not clear that this is the case, for it would seem that structure 2 would also adequately handle the semantics; that is, gunu tiko *be drinking* modified by gā *just, anyway.* Some more problematic examples are the following:

(40) e rua tiko gā na vuaka
it two stay just c-art. *pig*
There are just two pigs.

(41) ni rau bure vata tiko gā
that they-2 dwelling together stay just
that they were lodged in the same dwelling

(42) e na siga tiko gā oyā
on c-art. *day stay just that*
still on that day

(43) e kilai tiko gā me o Sē
he known stay just that p-art. *Se*
who was simply known as Se

In example (40) the *still, keep on* reading is not present at all. Thus it might be claimed that structure 2 is necessary to account for this kind of reading and structure 3 for (39a). However, example (40) might also be explained in that quite a number of special rules govern quantifiers in Fijian.

The semantics of (41), (42) and (43), however, cannot be explained away so easily. The *still, keep on* reading is absent from (41) and (43), but decidedly present in (42). Although further research is needed to put the matter beyond doubt, it would seem it is not a matter of one of the structures applying uniformly, but that now one, now the other applies in different circumstances. These circumstances, of course, need further delineation, perhaps in terms of phrase type or co-occurring word classes.

4.13 The combination mada gā is also problematic.

(44) monī curu saka mada gā mai i loma
that-you enter sir please just hither to inside
Please come in anyway, sir.

It is again tempting here to propose structure 3 as the operative one (curu modified by mada gā as a unit). However, it seems to me that mada may be operating according to structure 1; that is, gā modifies the verb first and then mada modifies the whole. Indeed, in example (44) the omission of mada does not change the meaning substantially except to be more polite. Structure 1, then, is the analysis I am tentatively suggesting, but there are more difficult examples and it may be the case that, as suggested for slot 46 in 4.12 above, we have one structure being employed on some occasions, another structure on others.

Clearly what is required is a much deeper syntactic analysis of the various relationships. I am simply mentioning some of these structural problems briefly just to give some idea of the nature of the problems and the complexity involved.

4.14 This complexity becomes even more elaborate when a number of the markers listed in 4.9 occur together.

(45) ni Ø sā kasou rawa sara tū mada gā na tūraga
that he now drunk get very stand mind just c-art. *man*
that the man had become absolutely drunk

In examples such as this, which of the preceding markers, if any, is gā in construction with? If my tentative analysis for mada is correct (cf. 4.13 above), we need not consider mada further.

Regarding tiko, toka, etc. (slot 46), gā is never in direct construction with these elements if wale, sara, or tale is also present. In such cases, the gā always modifies this element and never slot 46. Thus wale tiko gā breaks down into wale gā plus tiko, sara tū gā into sara gā plus tū, tale toka gā into tale gā plus toka, and so on. Where these three elements are absent however, it may be that slot 46 does enter into relationship with gā on occasion according to structure 3 (cf. 4.12).

Regarding kece and dina, if any marker comes between them and gā, the structure 3 relationship spoken of in 4.10 and 4.11 can no longer apply.[31]

We are left then with the situation that whenever one of the markers wale, sara, and tale occurs in a group of markers that also contain gā, it is they that enter into a structure 3 relationship with gā and not one of the other markers. When any two or all three of these markers co-occur, the rules are as follows.

In the combination wale sara gā, sara and gā enter into construction first, then modify wale and the whole combination modifies the verb. In the combination wale tale gā, gā often modifies both wale and tale at the same time; that is, we have a wale gā and a tale gā in operation together. Sometimes this provides ambiguity.[32]

(46) au raica wale tale gā na ikarua
I see-it only again just c-art. *second*
I too saw only the second one. OR *Only I saw the second one as well.*

In the combination sara tale gā, tale gā are always in construction as a unit (structure 3). Whether sara, like wale above, is also in construction with gā, I am not sure. I am inclined to think that it is not, but possibly its behaviour varies. In the combination wale sara tale gā, we actually have a wale sara gā (treated above) combined with a tale gā.

4.15 It was mentioned in 4.14 above that sara gā modifies wale when they occur together. In fact both sara and sara gā can modify the verb itself (slot 22), the modifying verb (slot 25), the adverb (slot 26), or any one of a number of markers, namely lō (slot 28), yādudua, etc. (slot 29), cake and sobu (slot 33), makawa (slot 37), kece, etc. (slot 39), wale (slot 41), and dina (slot 42). When, as is of course often the case, a number of these elements occur together, sara (gā) modifies the one closest to it.[33]

(47)a. e cakava sara o Viva
she do-it immediately p-art. *Viva*
Viva did it immediately.

b. e cakava (vaka)vinaka sara o Viva
she do-it well very p-art. *Viva*
Viva did it very well.

Somewhat surprisingly perhaps, if the adverb (vakavinaka in (47b)) comes after the sara (gā) instead of before (that is, if it is filling slot 58 instead of slot 26), although the reading may revert to the expected (?) *John did it immediately, well*, the sentence is in fact more likely to be read as synonymous with (47b), the adverb and sara (gā) being still therefore in construction.

4.16 When cake or sobu (slot 33) occur after the adverb of slot 26, the combination is interpreted in a comparative sense.[34] Thus if cake is substituted for sara in example (47b), the meaning is *Viva did it better.* This reading is definitely not possible, however, when the adverb is filling slot 58. Note that if sara occurs straight after the adverb plus cake (or sobu), it modifies that combination: vakavinaka cake sara *much better.*[35]

4.17 The marker dina too (slot 42) can optionally modify the adverb and some of the preceding markers, namely lō (slot 28), makawa (slot 37) and kece, etc. (slot 39).

4.18 The marker toka (slot 46), though frequently having the meaning already provided in 1.1, is also used somewhat idiomatically to provide a meaning of *somewhat, quite, rather, moderately.*

(48) Ø sā vinaka toka
it now good moderately
It's all right.

When used after an adverb of slot 26, toka is often to be understood in the same sense. Thus, if it is substituted for sara in example (47b), the reading is *Viva did it okay, Viva did it moderately well.* The interpretation of toka in this way depends largely on the word class of the element it is modifying.

4.19 gā, wale gā, and perhaps also mada gā, can modify at least some adverbs in the same way that sara (gā) can modify any adverb, regardless of whether it is filling slot 26 or 58.

(49) era sā mani druka gā vakarua
they now so lose just twice
So they lost only twice.

4.20 gā, wale gā, sara gā, and wale sara gā can also modify oqō, etc. (slot 56), which are of course also fulfilling an adverbial function in their time or place reading.[36]

(50) keimami raici koya wale gā oqō
we-excl. see him only just now
We saw him just now.

5. MARKERS IN THE NOUN PHRASE

Though this paper is concerned with the verb phrase, treatment of the verb phrase markers would be very incomplete without some reference to how these same markers occur in noun phrases. Uses in the verb phrase or in the noun phrase are often very closely related.

5.1 Nominalisations

When a verb phrase is nominalised, the conjunctions (slots 1 and 2) cannot occur with the first verb phrase, though they can occur (except for ia) with later verb phrases in a complex nominalisation. Nominalisations are introduced by the common article na. This is followed by the no- or ke- form, as appropriate

(cf. Arms 1974:97-102), of the possessive pronoun which replaces the normal subject pronoun form. The remainder of the verb phrase remains exactly the same under nominalisation. All markers may occur except for tei, no doubt because of its exclusively imperative (though polite) force.

(51) na nodra sā baci curu koso e loma ni bose
c-art. *their now again enter across in middle of meeting*
their barging in and interrupting the meeting again

5.2 Relative nominalisations

Another construction which is a kind of nominalisation is illustrated in (52a) below:

(52)a. o ira na ā curu lō mai
p-art. *they* c-art. past *enter secretly hither*

b. o ira era ā curu lō mai
p-art. *they they* past *enter secretly hither*
those who had entered secretly

The relative nominalisation (introduced by na in (52a)) resembles closely the normal relative clause (introduced by era in (52b)). The antecedent of the relative nominalisation is almost always pronominal, either a pronoun in a proper phrase (as in (52a)) or the subject or object pronoun of slots 3 and 24.

This kind of nominalisation is identical in form with the above (5.1) except that there is no pronoun form serving as subject after the na.[37] tei again cannot occur. In addition, it is not possible to have the future na (slot 6) occurring next to the common article na, no doubt for phonological reasons (cf. Geraghty 1976:516).

5.3 Markers before the noun

Although in nominalisations almost the complete inventory of markers may occur, there are very few such markers that can occur before a 'true' noun - indeed only two: dui (slot 15) and vei- (slot 19).[38] In addition lewe (or lē) (slot 21) can occur with quantifiers in the noun slot, and yā- (also slot 21) with numbers (including vica *how many, few*).[39] One special marker, ika- 'ordinal', occurs only in noun phrases and only with numbers (including vica).

dui does not co-occur with the above fillers of slot 21, nor does vei-, except with yā-. In fact yā- must have vei- present in order to occur in this position in the noun phrase.

5.4 Markers after the noun

After the noun in the noun phrase, one or more modifiers may occur. These may be nouns, or verb forms serving an adjectival or participial function.

After the modifier, if any, the last elements[40] that may be part of the noun phrase are certain markers. The markers that can occur in this position in the noun phrase are fewer than those that can occur after the verb. There is a sizeable range of them nevertheless, and they are most easily described by reference to the list of markers already given in 1.1. All of those markers can

occur except for lesu, lō, kaya, laivi, cake, sobu, koso, oti, rawa, mai, yani, kina and vuā (slots 27, 28, 31 to 36, 54, 55, and 57).[41] The adverb slots also occur, but the fillers are more limited and have a more adjectival function. The markers and adverbs occurring in noun phrases are subject to the same ordering principles, etc. already expounded in the previous sections of this paper.

As is the case in verb phrases (cf. 4.15ff), some of the markers in noun phrases may enter into construction with the modifier before the two jointly modify the noun.

(53) na vatu lailai sara
c-art. *stone small very*
the very small stone

When a verb form is filling the modifier slot in the noun phrase, it is occasionally modified by a marker in the same way as it is in a verb phrase. In such a construction even the markers listed above as excluded from noun phrases may occur, with the exception only of the pronominal ones: kaya, kina and vuā (nos. 31, 55 and 57).

(54) na leqa tubu koso
c-art. *trouble grow across*
the trouble that had cropped up

This construction, however, does not occur all that freely.[42]

6. COMPLEX PHRASES AND CLAUSES

When a number of different phrases or clauses occur together, two different sorts of problem can arise. There can be a question as to where precisely the markers are to be placed, and also a question as to which element they are modifying. The two questions are of course often interrelated.

6.1 Noun phrases

Where two noun phrases occur in construction, a general process takes place which might be described as 'marker fronting'.[43] Marker fronting is a process by which, if a marker may occur in a number of positions in a phrase or clause complex without change of meaning, it will tend to take the earliest one. This is simply a tendency, not a rule, and can be countermanded by considerations of clarity, emphasis or style.

(55)a. o ira sara gā na cauravou
p-art. *they very just* c-art. *youth*

b. o ira na cauravou sara gā
the youths indeed

(56)a. nodra vale kece na marama
their house all c-art. *woman*
all the women's houses

b. nodra vale na marama kece
the house(s) of all the women

(57)a. na vale gona ni kana
c-art. *house hence of eat*
In the room for eating therefore ...

b. na vale ni kana gona
In the dining-room therefore ...

Example (55a) above is far preferable to (55b), which is close to being ungrammatical. Marker fronting would normally operate in examples such as this.

Example (56a) is ambiguous, as is the English gloss, meaning either *all the houses of the women* or *the houses of all the women* (or also *the house of all the women*). The kece can be interpreted as modifying nodra or vale. In spite of the fact that (56b) is unambiguous, the (56a) arrangement is frequently preferred for that reading, presumably because of the marker fronting tendency. Of course the resultant ambiguity often vanishes in context.

In examples such as (57), both arrangements are acceptable. In many such examples the arrangement of (57a) would again be the preferred one. However, in others the arrangement of (57b) represents a virtual (or actual) relexification and is equally acceptable. This is the case in the actual example cited (57a and b), as reflected also in the English translations.

However, it is not a matter of all markers behaving in the same way.

(58) na cava rī e ā kaya tiko?
c-art. *what indeed he* past *say-it stay*
What was he saying again?

In this example rī would normally front rather than follow tiko.[44] The marker tiko, on the other hand, cannot front as it is exclusively modifying kaya, the verb. Similarly, if in example (56) above the marker beka *perhaps* is substituted for kece, the resultant (56a) would be very acceptable, but (56b) would be decidedly rare and almost ungrammatical. The reasons for these differences of behaviour derive from the different points at which the respective markers enter the structure. Column four of the list in 1.1 makes broad suggestions having reference to this, but, as mentioned earlier, much more research is necessary in this area of structural relationships.

6.2 Quantifiers with noun phrases

The grammar of quantifiers with noun phrases is somewhat complex.

(59)a. e rua na vale
There are two houses.

b. eratou tarā e rua na vale
they-few build-it it two c-art. *house*
They built two houses.

c. eratou tarā na vale e rua
They built the two houses.

We are mainly concerned here with (59a) and (59b).

In (59a) it seems clear that e is the third person unmarked subject pronoun, rua is filling the verb slot, and na vale is a subject noun phrase specifying e.

When we turn to (59b), it is tempting to regard the sequence e rua na vale as simply an embedding of (59a). In some sense perhaps it is, and I have glossed e accordingly. However, the e in this instance is more like an article, and the grammar relating to e rua na vale in (59b) is significantly different from that relating to it in (59a). Firstly, only the markers lewe (or lē) and yā- (including tauyā- and veiyā-) may occur before quantifiers with noun phrases (59b) whereas a fair range of the ordinary pre-verbal markers can occur with quantifiers in verb phrases (59a). Secondly, whereas the e in (59a) is changeable for person and number, the e of (59b) is not (cf. examples (60a,b and d) below). Finally, the article-like e with quantifiers in example (59b) must change to the 'true' common article na after prepositions (example (61)) and may do so even without a preposition before lewe (or lē) to provide a change of meaning (cf. example (60c)).[45]

(60)a. eratou lewe lima tiko
There are five (of them).

b. eratou tiko e lewe lima
they-few stay it person five
Five people are present.

c. eratou tiko na lewe lima
The five people are present.

d. keitou tiko e lewe lima
*we-few-*excl. *stay it person five*
Five of us are present.

(61) mai na rua na vale
at c-art. *two* c-art. *house*
at two houses

The full range of post-verbal markers can occur with quantifiers in verb phrases, with the exception of the numerative ones (slots 29 and 39). The markers which may occur after the quantifiers in noun phrases, however, are the same limited set presented in 5.4, with the additional exclusion of the numerative markers (slots 29 and 39) and makawa (slot 37). The markers, depending however on their particular function, usually occur after the quantifier and not after the following noun (cf. example (67) in 6.4 below).

6.3 Order of subject, verb and object

The verb phrase is the basic unit of the Fijian sentence. It can, and very often does, constitute a complete utterance on its own. Before proceeding further, a point that needs to be stressed is that the Fijian verb phrase as described contains a subject pronoun, verb and object, in that order (cf. 1. and Arms 1974:25). Although the object is optional and the subject may at times be realised as zero, still the basic order in the Fijian verb phrase is SVO, and it is *immutable*.

However, some noun phrases serve to specify the subject and object pronouns of the verb phrase (cf. Schütz 1981:204). These I will term respectively Subject Noun Phrases (SNP) and Object Noun Phrases (ONP). When we get both a subject noun phrase and an object noun phrase occurring with a verb phrase, we get the very opposite situation to that of the above paragraph. Far from being immutable, these three phrases can occur in all conceivable orders.

The classically given order is VP ONP SNP.[46] The order VP SNP ONP also occurs. It occurs more readily when the object in the VP is the third person unmarked pronoun than when it is any other pronoun. The order SNP VP ONP is common and, in main clauses, probably outstrips the 'classical' order in terms of frequency. The order ONP VP SNP is considerably less common. To have both noun phrases before the verb phrase is rare. The order ONP SNP VP is probably more acceptable on the whole than SNP ONP VP.[47]

When only one such noun phrase occurs with the verb phrase, again all orders are possible: VP SNP, VP ONP, SNP VP, and ONP VP, this last being the least common.[48]

It is not of course a matter of these various orders being randomly chosen. One or other order will be selected according to the exigencies of style, emphasis, clarity, textual organisation, etc.

6.4 Markers and the subject and object

Since, as we have just seen in 6.3 above, we always have a subject in the verb phrase and often have a subject noun phrase outside it - and since likewise we may have an object in the verb phrase and an object noun phrase outside it - markers modifying the subject or object have a choice of positions in which they could conceivably occur: either in the verb phrase or noun phrase and, if the latter, either before the verb phrase or after it.

(62)a. era na lako kece yani o ira na kai Toga
they future *go* *all* *thither* p-art. *they* c-art. *inhabitant* *Tonga*

b. era na lako yani o ira kece na kai Toga

c. o ira na kai Toga era na lako kece yani

d. o ira kece na kai Toga era na lako yani
All the Tongans will go there.

All these orderings are acceptable. Examples like (62b), however, have to be treated with some caution, for as mentioned in 6.1 above and for the same reasons, if certain other markers (e.g. beka *perhaps*) are substituted for kece in (62b), it becomes rare and borders on the ungrammatical. Note however that (62c and d) are both very acceptable. Marker fronting does not in the case of some markers push strongly that they move forward outside the verb phrase. In the case of others however it does, and (62d) then becomes the preferable arrangement (cf. the behaviour of rī in 6.1, example (58)).

When an object is present in the verb phrase, ambiguity may arise as to whether a particular marker is modifying the subject or object.

(63) keimami sā raici ira kece na vūlagi
we-excl. *now see* *them all* c-art. *visitor*
We have seen all the visitors. OR *We have all seen the visitors.*

Example (63) is potentially ambiguous as between the two readings given. Context would in fact often make clear which reading is meant in examples such as this. However, in the absence of an overriding context, the more natural reading would be to interpret kece as relating to the object, no doubt because of its physical proximity. This is especially true when full pronoun forms occur. In the case of the third person unmarked pronoun however, there is no preference for the object and the ambiguity is resolved solely by context.

(64) era kauta kece mai na ivola
they carry-it all hither c-art. *book*
They brought all the books. OR *They all brought the books.*

Of course such ambiguities do not arise with kece when the subject or object is clearly singular or dual.

Ambiguities such as the above can be avoided by shifting the marker from the verb phrase to the noun phrase. Thus example (65) provides just one of the readings of example (63).

(65) keimami sā raici ira na vūlagi o keimami kece
we-excl. *now see them* c-art. *visitor* p-art. *we*-excl. *all*
We have all seen the visitors.

Note that this example is of the same basic type as (62b), but marker fronting is suppressed here in favour of clarity. The arrangement of (62d) might also of course have been chosen.

Some markers or marker combinations, e.g. gā, wale gā, tale gā, can provide three-way ambiguities in the verb phrase.

(66) au raici koya wale gā
I see him only just
I only SAW him. OR *I saw only him.* OR *Only I saw him.*

This last reading is more difficult to get outside a suitable context, but is possible nevertheless. It is achieved readily if the phrase o yau *I* is put after the verb phrase and appropriate intonation is added.[49]

When a verb phrase is followed by a quantifier with a noun phrase, the marker fronting tendency does not seem to operate. Thus (67a) and (67b) are both very acceptable.

(67)a. au raica gā e dua na tamata
I see-it just it one c-art. *person*

b. au raica e dua gā na tamata
I saw only one person.

However, if the noun phrase na tamata is dropped from example (67), marker fronting operates and the (67a) arrangement is preferred.

6.5 'Compressed' clauses

There are some clauses that enter into an especially close relationship with each other.

(68)a. e sega lī ni ra tukuna?
it not pray that they say-it

b. era sega lī ni tukuna?
Didn't they say it?

In (68a) the clauses are in such close relationship that lī is governing the whole utterance and can occur equally well after sega or after tukuna (this

is true also of (68b)). Any of the markers of slots 47 to 53 can behave in the same fashion.[50]

In (68b) the two clauses enter into an even closer relationship. Here the subject pronoun of the second (lower) clause has become subject of the first (higher) one and no subject pronoun occurs in the second clause at all (- not even a zero form is to be posited). In addition, of the pre-verbal markers, the tense markers, sā and rairai can only occur in the first clause; se,[51] qai, mani, baci and dau can occur in either, and the remainder can occur only in the second clause. Of the post-verbal markers, only those from slots 39 to 55 inclusive occur in the first clause. All can occur in the second. In some instances the markers occurring in the first clause appear semantically to have been brought forward (raised) from the second.[52]

The verbs that admit of this sort of 'compression' like sega are: kua (or kākua) *not* (imper.), rawa *be able* (if it is followed by ni *that*, but not me *that*), bera *not yet*. kua (or kākua) has the additional property of being able to occur with tei (slot 10) which sega can't. rawa can have bau and dui as additional pre-verbal markers occurring in either clause. bera can only occur in this way if preceded by the marker se *as yet* and no others (except for the conjunctions and pronouns). The se however can be dispensed with after the conjunction ni *when*.

7. CONCLUSION

In attempting to give an account of the surface structure of the Fijian verb phrase, I have gone into some detail as regards the markers and some of their properties inside, and even at times outside, the verb phrase. The markers are an extraordinarily rich and complex part of Fijian grammar. In treating their surface order I have hopefully been able to carry forward the sterling work of other grammarians in this frontier and wilderness area. I have though, as should be clear, only *scratched* the surface.

NOTES

1. Much of the basic research behind this paper was done in the early 1970s. I am especially grateful, however, to Ratu Sakiusa Veiwili Komaitai of Bau for his endless patience and willing assistance in my restudy of the material more recently. My thanks too to Tēvita Nawadra and Paul Geraghty of the Fijian Dictionary Project for helpful discussions when revising this paper for publication.

2. Including the subject pronoun in the verb phrase may seem unconventional in terms of some grammatical theories. The nomenclature is convenient for our purposes here, however, and is in current use (cf. Arms 1974:25 and Schütz 1985:chapter 6) for Fijian.

3. Although for convenience I have used the term 'object' regarding the indefinite construction, it is not my intention in this paper to tip the scales one way or the other regarding the controverted question as to whether the construction is to be viewed as transitive or not (cf. Arms 1974:60-69).

4. vakatekivū is not readily accepted by some speakers as a marker and is in fact infrequently found in that capacity.

5. For more detail on these prefixes, see 3.12.

6. The forms yāyādua, yāyārua, etc. and lēlēyārua, lēlēyātolu, etc. are also found occasionally in this slot with the same meaning. The simple forms yādua, yārua, etc. are not usually employed in this slot in verb phrases (except for yāvica *so many*) but occur freely in this slot, or perhaps slot 39, in noun phrases, cf. 5.4.

7. It seems to have been assumed in the past (e.g. Milner 1972:117) that sō and soti as verbal markers are free variants. This is not so. soti is the more usually encountered form, and sō may only be substituted for it after the negative verbs segā and kua (or kākua) and the question words cava *what*, cei *who*, and vei *where*.

8. The first meaning given for each of the members of slot 46 is the meaning the form has when used as a main verb; the second, that which it has as a marker. Nevertheless, shades of the main verb meaning remain with the markers (especially voli). Because of this and because they are shorter, I will use the main verb meaning when glossing these markers in examples.

9. Although lā is good Bauan, it is rarely enough used in Standard Fijian. I have glossed gā and lā identically but some informants claim that lā conveys more of a sense of nonchalance or informality. This is supported by the fact that its typical use in written Standard Fijian is in direct speech. It evidently gives a colloquial or regional flavour to the discourse.

10. In conversational Standard Fijian, especially among non-Fijians and Fijians speaking in mixed racial company, gona can replace oyā in any situation. Hence I include it in slot 55. However, it would not as a rule be used in this way in the written language or in 'good' spoken Fijian, but would only be used as in slot 50.

11. Milner in fact termed most of them "particles". Since Schütz (1985:chapter 36) employs this term differently, I am using Schütz's term "markers" to run less risk of confusion.

12. Some of the markers are given in column three rather than column two for reasons of space.

13. One could argue that adverbs too are markers, but I find it useful here to keep a terminological distinction.

14. When inversion takes place, the a drops from the pronoun au *I* and the e drops from the first person inclusive and third person pronouns. Sometimes, however, the e remains in its original position. Thus, era sā lako *they have gone* can under inversion become sa ra lako or e sa ra lako. This does give the appearance of e being interpreted as a conjunctive or modal element (indicative).

15. Note that na (slot 6) is not subjected to inversion at all except in these particular cases. Although rairai (slot 7) is not listed here as subject to inversion, it will be seen in 2.2 that it has some particular properties that achieve a similar effect.

16. The individual glosses proper to tale and gā are given here, even though when occurring together they take on a special semantics, *again*, cf. 4. and note 27.

17. The abbreviations c-art. and p-art. occurring in glosses stand for 'common article' and 'proper article' respectively.

18. It is in fact not just slots 22 to 24 that can be repeated for compound or conjoined verbs, but the prefix slots (nos.19 to 21) as well. Indeed the slots may occur on one verb but not on the other. Note the vaka- form in example (6).

19. I am grateful to Paul Geraghty (personal communication) for pointing out that ka is a feature of the written language rather than the spoken, not just here but generally.

20. The symbol Ø has been introduced into this and some other examples to represent a zero-form pronoun.

21. se *or* can also conjoin nouns in this way. If it is employed in examples like (13), the common article na drops out.

22. kaya is a composite form representing kei plus the third person unmarked pronoun. The forms vuā and kina (slots 57 and 55) are similar composite forms, cf. Geraghty 1976:510.

23. A type of relative clause that does not appear to have surfaced in the grammars is one introduced by me. In such a case ka may never accompany me to introduce the clause.

(69) erau ā vakadrukai au ruarua ka sega na kā me u
they-two past *beat me both and not* c-art. *thing that I*

na sauma rawa kina vei rau
future *answer-it able to-it to them-two*
They had both beaten me and there wasn't a thing I could answer them.

24. Just as kē vakā is synonymous with kē, so also a special combination dē dua is synonymous with dē in the meaning *perhaps*.

25. Notice how the se me combination is functioning much as it did in example (39).

Note too that a second kē (which can be translated as *then*) is put introducing the main clause when the condition referred to is now realised and cannot therefore be changed (cf. example (31)). (In such cases the conditional clause always precedes the main clause.) Where the condition is not yet realised, kē cannot introduce the main clause.

A very unusual feature in conditionals when me is employed is that the negative is not necessarily kua (or kākua) as is usually the case after me, but may be sega.

(70) kē vakā me ā sega (or kua) ni lako mai, kē ...
if like-it that past *not that go hither then ...*
If he hadn't come, then ....

There is in fact a slight difference in meaning depending on which negative is used. In the sega version, the sentence is synonymous with the same sentence without me in its contrary to fact interpretation. In the kua version, however, there is an implication that the actor might, or should, have restrained himself. A more precise translation of this version might be: *if he had refrained from coming, then ....*

26. Like some of the other prefixes of this slot, nā- and tā- occur only with a rather limited number of roots. Their meaning would appear to be *turn* and *around, habitually, be in special state of* respectively. The prefix

sau is even more restricted and I am unsure of its semantics; occurring usually preceding tā-, it seems to reinforce that meaning.

27. One clear exception to this statement would be tale gā *also* (cf. 4.9). Although the meaning of this combination is related to that of its component parts tale *again* and gā *just*, it would appear nevertheless to be idiomatic.

28. In fact in example (38) the gā could be in combination with the preceding markers or with rua, or possibly even with both, cf. also 6.4.

29. gā may occur with the other members of slot 46 to provide a similar meaning of 'continuing', nuanced according to the marker in question.

30. kece gā does not actually occur much in the spoken language, kece sara or kece sara gā being preferred. It is common, though, in the written word, especially in religious texts.

31. This is further evidence that kece gā and dina gā are weak or even marginal combinations, cf. note 30 and 4.11.

32. It will be noted that wale gā and tale gā here are modifying pronominal elements. This phenomenon and the possible disambiguation of such sentences will be treated in 6.4.

    In contrast to example (46), wale tale gā sometimes breaks down into wale in the sense of *in vain, fruitless(ly)* on its own, and tale gā.

    (71) e sā lesu wale tale gā mai o Pita
    he now go-back fruitless again just hither p-art. *Peter*
    Peter too has come back empty-handed.

33. In the case of slots 28, 29 and 33, however, (lō, yādudua, etc., cake and sobu), sara (gā) sometimes ignores them, as it were, and modifies the preceding adverb or verb.

34. sobu occurs with only a few words in this comparative sense, e.g. cā *bad*, lailai *small; few*, malumalumu *weak*.

35. One also encounters sequences such as (vaka)vinaka duadua (or more rarely (vaka)vinaka taudua) *best* where at first sight it might appear that duadua (or taudua) is modifying (vaka)vinaka just as cake does. However, this is hardly the case. The meaning seems adequately accounted for if duadua (or taudua) are seen as modifying the pronominal element. Thus, adapting example (47) again, what would be literally *he alone did it well* would be understood idiomatically as *he did it best*.

36. As indicated in 2.22, this oqō can also occur outside the verb phrase. When it does, the markers modifying it accompany it (cf. also 6.4).

    (72) wale ga oqō keimami raici koya kina
    only just now we-excl. *see him then*
    Just now we saw him.

37. The true subject is already clearly expressed in the antecedent.

38. dau has not been included here since its occurrences in noun phrases are adequately accounted for in terms of relative nominalisation (cf. 5.2), e.g. na dau qoli *the fishermen*. (This still applies even if one regards the combinations as somewhat idiomatic and writes them as units, na dauqoli.) dau can also occur with closely related meaning as a noun.

39. For more on the status of quantifiers, see 6.4. lewe can occur with closely related meaning as a noun.

40. The markers are the last units within the basic phrase. Of course relative clauses and phrases modifying the noun phrase may occur after the markers.

41. It could be argued from their ordering properties and semantics that some of the markers are no longer functioning as markers but as ordinary modifiers in the modifier slot for nouns. This seems likely for tani and makawa and for some uses of wale and dina.

42. One of the markers before the verb, namely dau, may also occur in this way, and does so frequently.

(73) o cei soti na tamata dau cudrucudru oqori?
p-art. *who so-much* c-art. *person* habit. *angry* *that*(2)
Who are those hot-tempered people?

43. The term "fronting" is used by Paul Geraghty, to whom I am grateful for helpful observations on this aspect of Fijian grammar. Any errors of fact or analysis in this paper, however, are of course entirely my own.

44. A factor here too is the presence of a question word. Some special provisions can apply to these, cf. also example (75) in note 47.
Note that the marker rī is seeking (or expressing) confirmation of what should be (or is) known already. This idea is best translated in example (58) by *again*, which is here seeking a repeat of information.

45. This meaning change in terms of definiteness can be achieved generally by placing the quantifier after a noun phrase, cf. example (59c). If yā- occurs, the e drops and the quantifier fills slot 29 (cf. note 6). It would seem that it may fill that slot on other occasions too (and if lewe or lē occurs the e may optionally drop), but more generally the quantifier comes at the very end of the noun phrase.

46. This and the following statements really need accompanying discussion and evidence, but to provide it here would take us too far afield from the topic at hand.

47. These two orders were erroneously declared impossible in Arms 1974:26. Here, however, are examples:

(74) o iratou na "All Blacks" o keitou na timi
p-art. *they-few* c-art. *All Blacks* p-art. *we*-excl.-*few* c-art. *team*
ni Viti keitou na vakadrukai iratou sara gā edaidai
of Fiji we-excl.-*few* future *beat* *they-few very just today*
We Fijians will really beat the All Blacks today.

(75) na tamata oyā na cava e cakava tiko?
c-art. *person that* c-art. *what he do-it* *stay*
What's that person doing?

It is the presence of the question word cava that makes example (75) sound particularly reasonable. Such question words usually go before the verb phrase anyway, so the fronting of the noun phrase as well does not make the sentence sound convoluted.

48. Other noun phrases too - introduced by prepositions usually - may occur in virtually any order as amongst themselves or in relation to the subject and object noun phrases, but these are not of immediate concern here.

49. A lot more, of course, could be said on changes of meaning and disambiguation brought about by intonation.

50. The placement of some of these markers may depend on what precisely they are modifying.

51. se can in fact occur in both clauses at once, cf. Arms 1974:1269.

52. In many respects the two clauses look to be collapsing into one, with sega *not* becoming more of a marker than a verb. This is precisely the situation in a number of dialects. The corresponding negative words are simply markers.

REFERENCES

ARMS, David G.

1974 Transitivity in Standard Fijian. Doctoral dissertation, University of Michigan.

1978 Fijian "sā" and "se" aspect. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, 1241-1272. *PL*, C-61.

CHURCHWARD, C. Maxwell

1941 *A new Fijian grammar*. Sydney: Australasian Medical Publishing Company.

GERAGHTY, Paul

1976 Fijian prepositions. *The Journal of the Polynesian Society* 85:507-520.

HAZLEWOOD, David

1850 *A compendious grammar of the Feejeean language ....* Vewa, Feejee: Wesleyan Mission Press.

MILNER, George B.

1972 *Fijian grammar*. Suva, Fiji: Government Press.

SCHÜTZ, Albert J.

1981 Specification as a grammatical category of Fijian. In Jim Hollyman and Andrew Pawley, eds *Studies in Pacific languages and cultures in honour of Bruce Biggs*, 197-207. Auckland: Linguistic Society of New Zealand.

1985 *The Fijian language*. (Manuscript, to be published shortly.)

Arms, D.G. "Surface order in the Standard Fijian verb phrase". In Geraghty, P., Carrington, L. and Wurm, S.A. editors, *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*.
C-93:199-230. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.199*

CLASSIFICATION OF NIUEAN VERBS: NOTES ON CASE

Juliette Levin and Diane Massam

0.0 INTRODUCTION

In this paper we present a classification of Niuean verbs compiled from McEwen's (1970) *Niue dictionary* and various texts.[1] The aim of this classification is to shed light on properties of surface-ergative Case systems, such as that exhibited by Niuean. Though our findings are preliminary in many ways, interesting dependencies between thematic structure, grammatical relations and Case properties of verbs are revealed.

Verbs taking sentential complements are seen to fall into two basic classes: those with ergative subjects and those with absolutive subjects. Ergative Case-marking on subjects of verbs taking sentential complements appears to be counter-evidence to Safir's (1982) proposal that sentences (S's) cannot receive Case.[2] Whether directly or indirectly, such sentential complements must be viewed as receiving absolutive Case. Of interest is the fact that such verbs are all able to take NPs as direct objects as well. Absolutive Case-marking on subjects of verbs taking sentential complements argues for a theory in which sentences may, but need not, receive Case. The fact that the absolutive subject class includes all Raising-to-subject verbs in Niuean provides further empirical support for the second half of "Burzio's generalisation" (Burzio 1981). This generalisation states that: "A thematic role can be assigned to the subject if and only if Case is assigned by a verb to its object." As we will see, the Niuean data argues against T → A, while strengthening the claim that A → T. Such data are also compatible with Perlmutter's (1978) "Unaccusative Hypothesis".

Classification of Niuean verbs has also allowed us to examine data concerning the status of passive structures in Niuean. As discussed in Chung 1978, certain verbs in Niuean (and the entire Tongic subgroup) show evidence of the Proto-Polynesian passive suffix *-Cia. In Niuean, we find that some such verbs are syntactically intransitive and have lexically determined 'passive' interpretations, while others function as regular syntactic transitives. Other structures with optionally expressed agents include a class of verbs marked by the prefix ma-. These verbs form a fairly unified semantic class, and exhibit 'passive' interpretation, as well as syntactic intransitivity. Finally, certain canonically transitive verbs may be given passive interpretations when they appear with single absolutive arguments, despite the absence of any passive morphology. Such constructions indicate that the historical loss of passive is not yet complete, and that the presence or absence of an ergative-NP can determine active or passive voice respectively.

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 231-244. *Pacific Linguistics*, C-93, 1986.

1.0 BASIC CASE ASSIGNMENT

Before examining verb classes in Niuean and their significance for Case Theory, it will be helpful to outline the Case theory and mechanisms for Case assignment adopted in this paper.

We follow Marantz 1984 in classifying languages with surface-ergative Case systems together with nominative-accusative languages. True (or deep) ergative languages differ from nominative-accusative languages and surface-ergative systems in terms of the underlying correspondences between semantic roles and grammatical relations. The Nominative-Accusative/Ergative opposition proposed by Marantz is found in (1).[3]

(1) Semantic roles and grammatical relations (Marantz 1984:198)

Roles	Nominative-Accusative (includes surface-Erg.)	(Deep) Ergative
AGENT	SUB of [+trans], [+log sub] verb	OBJ of [+trans], [+log sub] verb
PATIENT/THEME	OBJ of [+trans], [+log sub] verb	SUB of [+trans], [+log sub] verb

Though Niuean has surface-Ergative Case-marking, it is clearly not an Ergative language as defined in (1), as we can see from the structure of the simple sentences in (2):

(2)a. Ne fakifaki e ia e fua moli. (M-29)
Pst *pluck* Erg *he* Abs *fruit orange*
He plucked an orange.

b. Ne hapo he tama e fuapolo. (M-77)
Pst *catch* Erg *child* Abs *ball*
The boy caught the ball.

Niuean is a VSO language with strict word-order.[4] In (2), the SUB of each sentence has the semantic role of AGENT, while the OBJ has that of PATIENT/THEME, just as in the English glosses provided. Thus, in Marantz's system, both Niuean and English are classified as nominative-accusative languages. What we must now ask is how nominative-accusative Case systems differ from surface-ergative Case systems.

This can be captured by the correspondences between grammatical relations and Case-marking shown in (3):

(3) Grammatical relations and Case-marking in NOM/ACC languages

Grammatical relation	Nom/Acc	(surface) Erg
A. SUB of [-trans] V	NOM	NOM (ABS)
B. SUB of [+trans] V	NOM	ACC (ERG)
C. OBJ of [+trans] V	ACC	NOM (ABS)

We will continue to refer to absolutive and ergative Case in Niuean, however, it should be clear from (2) and (3) that we are still speaking of a NOM/ACC system in terms of grammatical relations.

What remains to be formalised is the difference of Case-assignment which results in the two surface patterns shown in (3). The Case Filter (Chomsky 1981) given in (4) requires that every NP have Case, where Case may or may not be spelled out morphologically.

(4) Case Filter (Chomsky 1981)
*NP where NP has a phonetic matrix but no Case.

(5) The Case Marking Principle (Marantz 1984)
Case is determined under government/Government.[5]

We also adopt the Case Marking Principle in (5), where V is said to govern NP/VP and VP governs NP/S. In nominative-accusative languages, verbs assign accusative Case to their objects, and VPs assign nominative Case to their subjects. In ergative systems however, verbs regularly assign absolutive (=nominative) to their objects under government. Subjects governed by VP show up as absolutive (=nominative) with intransitive verbs, and as ergative (=accusative) with transitive verbs. These two systems can be seen to differ minimally if we allow both V and VP to determine nominative Case cross-linguistically. All that need be said is that assignment of nominative Case is *obligatory*. In so-called accusative languages, VP determines nominative, while in surface-ergative systems, it is the verb which determines nominative. If nominative Case cannot be assigned by V in such a system (i.e. if there is no NP/VP) then nominative (absolutive) percolates to the VP projection and is assigned by VP to the NP it governs, namely the subject. In accusative systems where VP cannot assign nominative case (i.e. in so-called "unaccusative constructions"), nominative can percolate from VP down to V, where it can then be assigned to NP/VP. We formalise this mechanism as follows:

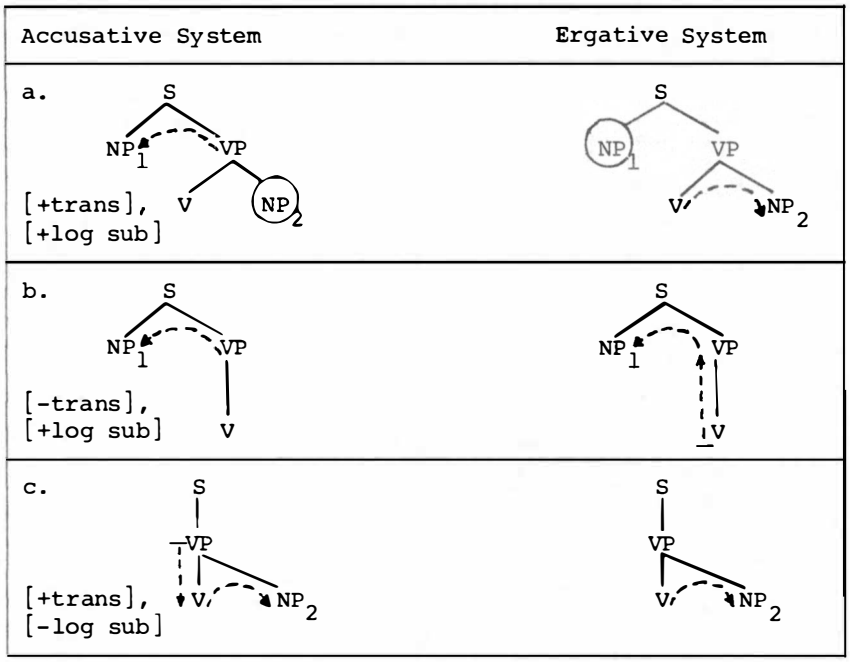
(6) Conditions on Nominative (=Absolutive) Case (CNC)

A. Nominative Case must be assigned wherever possible.

B. Nominative Case determined by X^n is transferable within X^{max}.

We can now evaluate the formal differences between accusative and surface-ergative systems within this framework, as shown in (7), where arrows indicate Case assignment.

(7) Assignment of nominative (absolutive) Case



The circled NPs in (7) receive Accusative (ergative) Case via their governors. Thus, the single parameter which distinguishes accusative from surface-ergative Case assignment in this system, is the choice of which element determines nominative Case assignment. In surface-ergative systems it is the verb, while in nominative-accusative systems it is the verb phrase.

We are now in a position to evaluate verb classes in Niuean with respect to the Case-marking schema shown in (7). We will concern ourselves with two basic questions. First, given that the Case Filter (4) refers specifically to NPs, what are the Case relations in Niuean when an NP in (7) is replaced by a sentential argument? Second, passive is usually analysed as a mapping between structures (7a) and (7c), with subsequent change of grammatical relations, making NP_2 the surface subject of VP. If this is so, are there structures like (7c) in Niuean which have passive properties?

2.0 VERBS WITH SENTENTIAL ARGUMENTS

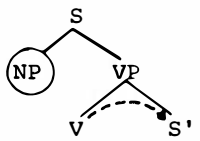
Given that Niuean exhibits fairly strict VSO word-order, it is difficult to distinguish the argument structures shown in (7b) from those in (7c). Nevertheless, we can make a primary descriptive division in verbs with sentential arguments by separating those with a single sentential argument from those with multiple arguments. We will first discuss the subdivisions of the class of verbs taking multiple arguments.

2.1 [+trans] verbs with S' complements

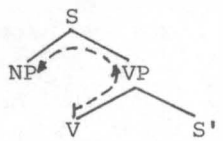
Verbs which assign syntactic roles have the feature [+transitive] as a lexical property (cf. Marantz 1984). Both fakifaki *pluck*, and hapo *catch*, in (2) are [+transitive], as they assign syntactic roles to their objects. As we saw in (7), absolutive (nominative) Case, in these instances, is assigned by V to the NP it governs, and ergative Case on the subject is determined by the VP. Notice however that replacing an NP-object by an S' leaves open the possibility that absolutive Case not be assigned by V, as long as it is eventually assigned to some argument. This is possible because the Case Filter (4) refers to NPs, not S's, thus allowing non-Case-marked S's. The two possibilities for absolutive Case-assignment are illustrated in (8):

(8) Absolutive (=nominative) Case Assignment with S' complements

a.



b.



If S's required Case, then (8a) would result, and all [+transitive] verbs with S' complements would surface with ergative subjects. On the other hand, if S's could not be Case-marked, then the CNC (6) would dictate the Case schema in (8b) and subjects in these constructions would be found Case-marked as absolutive.

The Niuean data proves interesting in that it supports an analysis where both possibilities in (8) are realised. That is, [+transitive] verbs with S' complements subdivide into two classes, one which appears with ergative subjects, and the other with absolutive subjects. In (9a,b) two Niuean sentences are given to illustrate the evidenced Case arrays of (8a,b) respectively.

(9)a. Kua iloa e mutolu ke mailonga e mahani he langi. (M-180)
Perf *know* Erg *you* Sbj *distinguish* Abs *signs of sky*
You know how to distinguish the appearance of the sky.

b. Kua lali a ia ke vangahau. (M-146)
Perf *try* Abs *he* Sbj *talk*
He is trying to talk.

The class of verbs which pattern with iloa *know* in (9a) is quite small. A list of these verbs, which are labelled TS-ERG, is given in (10).

(10) V erg-NP S': TS-ERG verbs

iloa *know, know how*; kamata *begin*; kitia *see*; manatu *think, wonder*; longona *hear, feel*; talahaua *say*.

The example in (9a) shows iloa taking a non-finite complement.[6] However, TS-ERG verbs can also occur with finite complements, as shown in (11).

(11)a. Kua iloa nī e au to tutupu e tau mena he pō ia. (S-125)
Perf *know* Emph Erg *I* Fut *grow* Abs Pl *thing on night that*
I just know that things (clouds) would gather that night.

b. Ne kitia he kau kaihā kua mate tuai e molī he fale. (S-126)
Pst *see* Erg *group thief* Perf *die* Perf Abs *lamp in house*
... the thieves saw that the lamp in the house had gone out.

Such data appears to be clear counterevidence to Safir's (1982) hypothesis that S's cannot receive Case. Though absolutive Case does not surface morphologically on S's, ergative Case on the subject indicates that absolutive has been assigned. Furthermore, given the CNC, absolutive Case must have been assigned, leading us to conclude that the S's in (9a) and (11) have all been Case-marked as absolutive. The fact that TS-ERG verbs show up with both finite and non-finite complements also argues against a weaker version of Safir's proposal, where the subset of tensed (or finite) S's could not be Case-marked.

One might argue that the S's in question are immediately dominated by NP, thus accounting for their absolutive Case-marking. In fact, all TS-ERG verbs listed in (11) can also take NP-objects. Examples with iloa *know* and kitia *see* are given in (12):

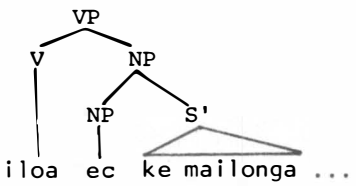
(12)a. Kua iloa tuai e lautolu oti a au. (S-248)
Perf *know* Perf Erg *they* *all* Abs *me*
All of them know me.

b. Kua kitia e maua e pusi haau i loto he tau fiti. (TM)
Perf *see* Erg 1.Ex.Du Abs *cat* *your* Loc *inside of* Pl *flower*
We see your cat among the flowers.

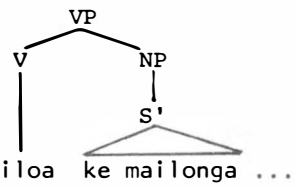
Arguing that such S' complements are in fact NPs with either of the headless relative structures shown in (13), would predict that extraction from such clauses would result in violations of the Complex NP Constraint (Ross 1967) or, in more recent terms, of subjacency (Chomsky 1981).

(13)(=9a) a.

(ec=empty category)



b.



However, extraction from such clauses is grammatical, as shown by the constructions in (14), where relativisation or ko-Clefting has occurred out of a complement S'.

(14)a. e ika ne iloa e koe ke hī
Abs *fish* Nft *know* Erg *you* Sbj *catch*
the fish which you know how to catch

b. ko Moka ne manatu e ia ke alofa
Pred *Moka* Nft *think* Erg *he* Sbj *love*
It's Moka that he thinks he loves.

Despite the fact that S' complements of TS-ERG verbs cannot be analysed as NPs, the correspondence between Case to NP objects and S' objects need not remain a coincidence. The subcategorisation frames for such verbs refer to NP or S'. If Case-marking is a lexically determined property of verbs, then TS-ERG verbs must be classified as Case Assigners in their lexical entries. We will call such verbs [+CA]. A verb which is [+CA] must assign Case to its internal argument(s). Such a hypothesis clearly needs to be tested on other languages with surface-ergative Case-marking where Case-marking to sentential complements is visible via ergative Case on the subject.

We now turn to a brief examination of [+transitive] verbs taking S' complements with absolutive subjects, such as lali *try* in (9b). A partial list of these verbs, which we will call TS-ABS verbs, is given in (15).

(15) V Abs-NP S' : TS-ABS verbs

amanaki *hope*; fakaanga *attempt*; foli *decide*; lali *try*; fakalata *think*; manako *want*; talifaki *expect*; amamanaki *hope*; piko *believe*.

The example in (9b) shows lali *try* taking a non-finite complement. Though the majority of verbs in this class are Control (Equi) verbs and as such take non-finite complements, TS-ABS verbs also appear with finite complements as the examples in (16) illustrate.

(16)a. Fakalata a Stan kua fakatau tuai he fifine e falaoa. (S-126)
think Abs *Stan* Perf *buy* Perf Erg *woman* Abs *bread*
Stan thinks that the woman bought the bread.

b. Piko e mangafaoa haaku ne fano a koe ki Sāmoa. (S-125)
believe Abs *family my* Pst *go* Abs *you to Samoa*
My family believed (mistakenly) that you were on your way to Samoa.

The class of TS-ABS verbs illustrates the Case-marking schema shown in (8b). Sentential objects, not being NPs, need not be Case-marked, and absolutive can be transmitted to VP and assigned to NP/S. The fact that such verbs do not assign Case to their internal arguments leads us to classify them as [-CA], that is, as non-Case-assigners.

Recall Burzio's generalisation, which is repeated in (17):

(17) Burzio's generalisation (Burzio 1981:170)

T ↔ A where: T = assignment of theta-role to subject
A = accusative Case assignment

In surface-ergative Case-systems like Niuean, A must refer to that Case governed by V, i.e. to absolutive. Given the two classes TS-ERG and TS-ABS, or [+CA] and [-CA], it is clear that -T ↔ -A does not hold. TS-ABS verbs do not assign absolutive Case to their S' complements, and yet a theta-role is assigned to the subject. It seems we have found the evidence Burzio himself hypothesised might exist in his discussion:

> ... our framework will not require that the statement [-A → -T;jl/dm] should hold for verbs appearing in other than the configuration in (138) [[NP]V[-A] ...NP, where NP/VP is governed by V and only by V;jl/dm]. For example we would expect that in a base form "NP V S" where there is no NP to assign Case to, the verb could very well lack the capability to assign accusative .... However, since we find no evidence that would ever falsify it, we will assume that (139) [-A → -T;jl/dm] holds categorically. (p.169)

The Niuean data is still compatible with the weaker form of (18) which says simply A → T.

Data on Raising-to-subject verbs also seems to lend support to this weakened well-formedness condition, or to its logical equivalent, -T → -A. Raising-to-subject verbs are a subclass of TS-ABS verbs, though they differ from the verbs

listed in (15) in that they are [-log sub], no theta-role being assigned to the position NP/S. A list of Raising-to-subject (RS) verbs is given in (18):

(18) Raising-to-subject verbs: V [$_{NP}$ec] S', V Abs-Np S'

maeke *can, be possible*; kamata *begin*; fakaai *not*; mahani *usual, customary*; teitei *almost*; fetamakina *nearly*.

In (19) we see three sentences related by the rule of Raising.

(19)a. Kua kamata ke hala he tama e akau. (S-158)
Perf *begin* Sbj *cut* Erg *child* Abs *tree*
The child has begun to cut down the tree.

b. Kua kamata e tama ke hala e akau.

c. Kua kamata e akau ke hala he tama.

In (19b) the lower subject is raised, while in (19c) it is the lower object which is raised to subject position.[7] As should be clear, Case of the raised NP is determined by the matrix clause, not the embedded clause. All RS verbs occur with absolutive subjects. The absence of RS verbs with ergative subjects is negative evidence supporting Burzio's proposal, -T → -A, or equivalently A → T.

In summary, data from transitive verbs taking S' complements provides evidence that S's may be Case-marked, though they need not be. Furthermore, we have seen that both TS-ABS verbs and RS verbs support only half of Burzio's generalisation, the half which states A → T, while the existence of a class of TS-ABS verbs in itself falsifies the stronger proposal T ↔ A, since sentential complements appear without Case, and a theta-role is assigned to NP/S.

2.2 Verbs with single S' arguments

As mentioned earlier, it is difficult to distinguish (7b) from (7c) in a surface-ergative VSO language. The same holds if NP in (7) is replaced by S'. S's do not take part in agreement nor in noun-incorporation, so it is difficult to test their status as surface subjects or objects. It so happens then that this particular class of verbs tells us little about Case-marking. A partial list of what we will call bare-S' verbs is given in (20) and several examples appear in (21).

(20) Bare-S' verbs: V S'

hangahanga *appear*; lata *be right*; lingalinga *probable, possible*; liu *again*; tonuhia *be right*; mitaki *good*; kelea *bad*.

(21)a. Kua lata ke fekapitingaaki a tautolu. (M-40)
Perf *right* Sbj *be-friendly*-Rcpr Abs *we*
It's right that we should be friendly with each other.

b. Kua kelea koa he pāki a koe. (S-129)
Perf *bad* Emph Caus *injured* Abs *you*
It's too bad that you were hurt.

(21) illustrates that both finite and non-finite clauses appear in this configuration. Bare-S' verbs which take ke-clauses, such as lata in (21a), must be

analysed as taking only a single external argument, since raising-to-subject cannot take place with these verbs. Assuming the S' to be in subject position in (21a), raising is not possible, since there is no empty position for the NP to move into. The similar semantics of verbs like maeke *possible, be able* and lingalinga *possible, probable* make it appear likely that argument structure (internal vs. external S' argument) is the essential factor distinguishing raising verbs from non-raising verbs.

3.0 PASSIVE AND CASE-MARKING

Our investigation of Case-marking properties in Niuean focuses on possible passive constructions for two main reasons. First, within certain theoretical frameworks, such as Government and Binding theory (Chomsky 1981), passive is characterised as the elimination of the Case-assigning properties of the verb. In our terms, we state that Passive changes a [+CA] verb into a [-CA] verb. That is, though in English passive eliminates accusative Case assignment to NP/VP, in Niuean it eliminates assignment of absolutive case to NP/VP. Second, Chung (1978) has argued convincingly that the surface-ergative Case system in Niuean is a result of a historical reanalysis of Proto-Polynesian passive constructions. The passive-to-ergative reanalysis is claimed to have done away with a syntactic passive altogether. We will investigate what a rule of passive would look like in a surface-ergative Case-system, given the mechanisms for nominative (absolutive) Case assignment presented earlier. The conditions of Nominative Case Assignment (CNC) stated in (6) together with the second part of Burzio's generalisation, actually predict the possibility of passive in surface-ergative languages without passive morphology, a possibility which is realised in Niuean.

3.1 Remnants of Proto-Polynesian passive suffix *-Cia

As discussed in Chung 1978, the remnants of Proto-Polynesian passive suffix *-Cia can be seen in several types of non-productive lexical derivational forms, some of which are verbs. Semantically, -Cia verbs can differ from their stems in terms of completion of an event, duration or lack of agency. Syntactically they are usually intransitive and select a subject corresponding to the direct object of the stem. The agent can be expressed by an oblique NP. What we would like to point out here is that in terms of syntactic argument structure, there is no formal way to distinguish a canonical intransitive verb from an intransitive verb with a -Cia suffix. Both positions can be controlled, and neither can be incorporated (noun-incorporation in Niuean is limited to non-subjects). The control facts are illustrated in (23):

(22)a. Ati kata vave e tama. (M-116)
and laugh soon Abs *boy*
and then the boy soon laughed.

b. Kua hulungia a Tapeu. (M-98)
Perf *lit up* Abs *Tapau*
Tapeu is lit up.

(23)a. Kua lali a ia ke vangahau. (M-146)
Perf *try* Abs *he* Sbj *talk*
He is trying to talk.

b. Kua lalilali a ia ke hoko ke ofania. (M-250)
Perf *persevere in trying* Abs *he* Sbj *arrive* Sbj *be loved*
He kept on trying to be loved.

Synchronically then, -Cia intransitive verbs are no different from other intransitive or stative verbs, aside from the optional appearance of an oblique agent with certain verbs.

Another class of verbs, those morphologically marked with the prefix ma-, have several features in common with the -Cia verbs just discussed. They are for the most part derived from syntactically transitive verbs and differ from their stems semantically in terms of completion and lack of agency. Some examples are given in (24).

(24) Ma-Verbs: V Abs-NP

mafuke	*opened*	fuke	*open*
mafuli	*overturned*	fuli	*turn over*
mahaku	*scratched*	haku	*scratch*
mahele	*cut*	hele	*cut*

The ma- verbs cover an interesting semantic class which includes verbs whose action results in an observable physical change of state on the patient. There are very few exceptions to this generalisation. Ma- verbs are distinct from -Cia passives in their dual function as adjectives. Thus, it appears that though verbs with either the -Cia suffix or the ma- prefix function as intransitive verbs whose sole arguments are semantic THEME/PATIENT, such morphology is limited to subsets of the lexicon. Neither process applies exclusively to [+CA] verbs, making them [-CA], thus disallowing either affix to be viewed as a true productive passive morpheme.

3.2 Passive without morphology

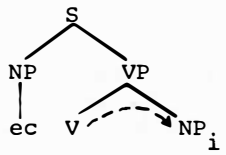
We now return to the characterisation of passive as the elimination of Case assigning properties of the verb. As mentioned earlier, within Government and Binding theory, passive morphology in English is argued to eliminate assignment of accusative Case by a verb to NP/VP. The deep object moves to subject position so as to avoid a violation of the Case Filter. Notice that within this theory, passive also involves a de-thematisation of subject position. Given the fact that the AGENT can be realised in a by-phrase in English, we will assume that assignment of theta-role to subject position is optional.

In trying to evaluate such a formal process for surface-ergative languages, we are faced with the same problems posed by Burzio's generalisation. That is, though we have set up equivalencies between accusative/ergative Case and nominative/absolutive Case, we do not want to claim that passive involves the dissolution of a verb's ability to assign ergative Case, since verbs never directly assign ergative Case. Rather, it is clear that the true nature of passive is to change verbs from [+CA] to [-CA]. Thus, in surface-ergative languages, passive should prohibit a verb from assigning absolutive Case. With V unable to assign Case, Case must be assigned by VP (CNC). As Case is assigned under government, this requires that an internal argument be externalised to receive Case. At this point, another question arises. Why is it that the NP in [V Abs-NP] cannot be interpreted as a VP-internal argument? Recall that Burzio's generalisation

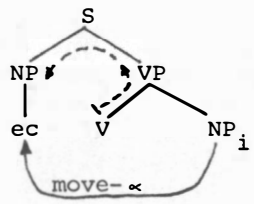
requires that in such a configuration, a theta-role be assigned to subject position. But if this is the case, then the passive interpretation will not be available.

Let us now review absolutive Case assignment as put forth in (7). In particular, we will examine (7c), which is repeated in (25a).

(25)(=7c) a.



b.



In (25a) the logical object of the verb is assigned absolutive Case, since absolutive (=nominative) must be assigned under the CNC, and NP_i must be assigned Case to satisfy the Case Filter. However, (6B), the second clause of the CNC allows Case to be transferred within the VP. This transference must occur just in case no theta-role is assigned to the subject, since otherwise, a violation of A → T results. Thus, as illustrated in (25b), absolutive Case may be transmitted from V to VP without any additional morphology. Should this occur, NP_i will not receive Case, and will be forced to move into subject position to receive Case under government by VP.

Such a process is formally identical to passive in nominative/accusative languages, except that instead of additional morphology changing a verb from [+CA] to [-CA], in surface-ergative languages the Case-marking properties of V are freely transferable to VP. Niuean appears to exploit this mechanism, to the point where one might speak of the (b) sentences in (26)-(28) as instances of passive without morphology.

(26)a. Ti nākai liu foki keli e au e tau mena momoui oti. (M-125)
and not resume kill Erg *I* Abs Pl *thing alive all*
And I will not again kill all living things.

b. Kua keli a ia. (M-21)
Perf *kill* Abs *he*
He is killed.

(27)a. Kua tāmate e laua a Mutalau. (M-307)
Perf *kill* Erg *they* Abs *Mutalau*
They killed Mutalau.

b. Ne tāmate a ia. (M-307)
Pst *kill* Abs *he*
He was killed.

(28)a. Afu he tau tangata e umu. (TM)
heap-up Erg Pl *man* Abs *oven*
The men heaped up the oven.

b. Afu e umu. (M-3)
heap-up Abs *oven*
An oven was heaped up.

Similar facts have been discussed by Tchekoff (1979) for Tongan, where a syntactically transitive verb appearing with a single absolutive NP can be interpreted in one of the three ways shown in (29), depending on the particular verb.

(29) Tongan: V[+trans] Abs-NP (Tchekoff 1979)

A.	kai *eat*	AGENT or THEME
B.	ʔave *bring*	THEME only
C.	tamate *kill*	AGENT only

In Niuean, certain verbs may be ambiguous as those in (29A), but no instances of (29C) have been found. In fact, (28b) shows the Niuean cognate, tāmate, with a (29B) reading. Further investigation into the status of ambiguous versus non-ambiguous interpretation of single absolutive arguments with [+transitive] verbs remains to be done. Even at this point however, the interpretation of [+trans] verbs with single absolutive NPs as passives, despite the lack of any overt morphology, lends support to the theory of Case proposed herein and, in particular, to the equivalency set up between nominative and absolutive Case, and to the Condition on Nominative Case proposed above. It was only within such a theory that the optional Case assignment to sentential complements could be formulated properly, strengthening A → T of Burzio's generalisation. It appears then that we have confirmation of a method of analysis which focuses on the integrity of Case-marking systems, both language-internally and cross-linguistically.

NOTES

1. Sources for example sentences are given in parentheses after each example. Seiter (1980) has been used as a secondary source, in addition to the primary sources listed in the bibliography. We would like to thank Jerry Malumaleuma for his work as a Niuean consultant.

2. Throughout we use Case to refer to abstract Case in the sense outlined in Chomsky 1981. Abstract Case is assigned by a verb to its direct object, by a VP (or INFL) to its subject, and by a preposition to its object. Abstract Case may or may not have a morphological realisation depending on the language and on the specific Case involved.

3. See B. Levin 1983 for a detailed discussion of Marantz's system of deep versus surface ergativity.

4. While the surface word-order of Niuean is VSO, we will follow Chung (1983) and Sproat (1985) in positing a VP at both D- and S-structure. Thus, the underlying word-order is SVO with a subsequent V-movement rule resulting in the surface VSO word-order. At the level where Case assignment is relevant, then, the word-order is SVO, as we have indicated in the tree-structures where Case assignment is illustrated.

5. See Chomsky 1981 and references therein and Marantz 1984 for a detailed exposition of the theory of government, and for consequences of the Case Marking Principle.

6. We refer to ke-clauses as non-finite, since they do not contain regular tense markers, and since their subjects can be controlled. They differ however from truly non-finite clauses, and are more like subjunctive clauses, in that they may appear with overt subjects, and in that it is possible to inflect ke for past tense (kua) if embedded under a past tense matrix verb.

7. This sentence illustrates raising from object to subject, a construction particular to Niuean among Polynesian languages, and problematic in many respects. For a discussion of this construction, see Levin and Massam 1984.

REFERENCES

BURZIO, L.

1981 Intransitive verbs and Italian auxiliaries. Ph.D. dissertation, MIT, Cambridge.

CHOMSKY, N.

1981 *Lectures on government and binding*. Dordrecht: Foris Books.

CHUNG, S.

1978 *Case marking and grammatical relations in Polynesian*. Austin: University of Texas Press.

1983 The ECP and government in Chamorro. In *Natural Language and Linguistic Theory* 1/2:207-244.

LEVIN, B.

1983 On the nature of ergativity. Ph.D. dissertation, MIT, Cambridge.

LEVIN, J. and D. MASSAM

in press Raising and binding in Niuean. In R. McGinn, ed. *Studies in Austronesian linguistics*. Athens: University of Ohio Press.

MARANTZ, A.

1984 *On the nature of grammatical relations*. Cambridge: MIT Press.

McEWEN, J.M.

1970 *Niue dictionary*. Wellington, N.Z.: Department of Maori and Island Affairs (M).

PERLMUTTER, D.M.

1978 Impersonal passives and the unaccusative hypothesis. In Jeri J. Jaeger et al, eds *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistics Society*, 157-189. Berkeley, CA: Berkeley Linguistics Society.

ROSS, J.R.

1967 *Constraints on variables in syntax*. Bloomington: Indiana University Linguistics Club.

SAFIR, K.

1982 Syntactic chains and the definiteness effect. Ph.D. dissertation, MIT, Cambridge.

SEITER, W.J.

1980 *Studies in Niuean syntax*. New York: Garland Press (S).

SPROAT, R.

1985 Welsh syntax and VSO structure. In *Natural Language and Linguistic Theory* 3:173-216.

STOWELL, T.

1982 Origins of phrase structure. Ph.D. dissertation, MIT, Cambridge.

TCHEKOFF, C.

1979 *La construction ergative en avar et en tongien.* Paris: Klincksieck.

Tohi tala ma e fanau lalahi Niue 1-3 (1960-63); *Lopinisoni kuluso* (1961); *Ko e ua e teau tau Hikihikifano* (1965); *Ko Pita ko e tangata hokatafua* and *tule mo Moka mo pusi ko kula* (TM) (1969). Wellington, N.Z.: Island Education Division of the New Zealand Department of Education for the Department of Island Territories.

Levin, J. and Massam, D. "Classification of Niuean verbs: notes on case". In Geraghty, P., Carrington, L. and Wurm, S.A. editors, *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics.* **C-93:231-244**. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.231*

WORD ORDER IN TUVALUAN

Niko Besnier

0. INTRODUCTION[1]

Traditionally, Polynesian languages have been viewed by typologists and descriptivists as prototypical VSO languages. It has been known for some time, however, that many Polynesian Outlier languages[2] exhibit non-verb-initial sentential constituent order configurations (hereafter referred to as word order), although little is known in detail about the syntax of these languages, let alone about their word-order mechanisms (Chung 1978 and Clark 1976 both devote a few pages each to a discussion of word order in Outlier Polynesian, and an unpublished paper by Reedy (1977) investigates the question for Takuu). More recently, the work of a number of researchers, such as Ochs (1982) and Duranti (1981) on Samoan, Hooper (1986) on Tokelauan, and Alexander (1981) on Rapanui, has shown that, even in non-Outlier Polynesian languages, the pragmatics of word order offers a much more complex picture than is commonly assumed. The question of the distributional patterns of word-order configurations in Polynesian languages in general, thus, deserves more attention.

This paper is an investigation of word order in the seven dialects of Tuvaluan, all of which share the same syntax (but differ in their morphology - Besnier 1986), characterised by a high degree of freedom in the order of sentential constituents. This study provides a functional explanation for the attested word-order variations on the one hand, and, on the other hand, for the fact that a number of logically possible word-order configurations do not occur.

In this paper, I shall show that, despite the word-order freedom exhibited by Tuvaluan, there is a basic order, and that this order is verb initial. It will be shown that the non-basic word-order variants can be explained functionally as encoding the pragmatic role of the nominal constituents of the sentence. Furthermore, word order interacts with case marking in transitive clauses in such a way that post-verbal, ergatively-marked agents are always marked for high agentivity; besides being dependent on the pragmatic structure of the clause, word order is thus also governed by semantic notions. This complex account, as will be seen, provides an explanation for the fact that a number of word-order configurations are not attested. Finally, I shall consider the Tuvaluan data in the light of the claim that word-order variation similar to that exhibited by Tuvaluan is typically symptomatic of change in process. I shall show that, contrary to this position, word-order variation in Tuvaluan appears to be a stable phenomenon. In conclusion, it will be proposed that the account of Tuvaluan word-order variation proposed here may be extended to at least some of the Outlier languages. Because strong genetic connections between Tuvaluan and some

---

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics,* 245-268. *Pacific Linguistics*, C-93, 1986.

Outlier languages have been suggested by Bayard (1967), Pawley (1967), and Howard (1981), special attention will be given in the last part of the paper to the relative roles of shared syntactic innovation and of independent developments to account for the word-order patterns encountered in the Polynesian language family.

1. THE DISTRIBUTION OF WORD-ORDER VARIANTS

In Tuvaluan, all constituent-order possibilities are grammatical, with the restriction that a transitive subject cannot precede the verb if no direct object follows the verb. Contrast, for example, the various grammatical word-order possibilities given the same set of three constituents in the transitive clauses in (1a-h) with the unattested variants in (1i-k), which are ungrammatical both with and without an ergative case-marker:[3]

(1) STV

a.	Ne ffuti nee Ioane te ika teelaa. Pst *pull* Erg *Ioane the fish that* *Ioane caught that fish.*	VSO
b.	Ne ffuti te ika teelaa nee Ioane.	VOS
c.	Te ika teelaa ne ffuti nee Ioane.	OVS
d.	Ioane ne ffuti te ika teelaa.	SVO
e.	Te ika teelaa ne ffuti.	OV (S deleted)
f.	Ne ffuti te ika teelaa.	VO (S deleted)
g.	Ne ffuti nee Ioane.	VS (O deleted)
h.	Ne ffuti.	V (S and O deleted)
i.	*(nee) Ioane ne ffuti.	*SV (O deleted)
j.	*(nee) Ioane te ika teelaa ne ffuti.	*SOV
k.	*Te ika teelaa (nee) Ioane ne ffuti.	*OSV

Example (2) below illustrates the word-order possibilities for an intransitive sentence:

(2) STV

a.	Koo fano Ioane. Inc *go* *Ioane* *Ioane has left.*	VS
b.	Ioane koo fano.	SV
c.	Koo fano.	V (S deleted)

Several types of oblique noun phrases, such as temporal and locative expressions, also participate in this scrambling effect; others, such as comitative noun phrases and middle objects, which are shown elsewhere to be treated by Tuvaluan syntax as oblique noun phrases (Besnier 1981a, 1986), are fairly fixed in a post-verbal position. This discussion, however, will be restricted to the relative position of subjects, verbs, and direct objects.

The various word-order configurations are associated with different case-marking strategies, a fact that will prove important in this discussion. The facts can be summarised as follows: post-verbal transitive subjects are always marked with an ergative preposition nee (or e in the three Northern dialects), while all other noun phrases are either unmarked for case or marked by an absolutive/neutral preposition a (this morpheme, which is reviewed briefly in Wang

1976, appears to play many roles, one of which is the marking of contrastiveness). The ergative preposition nee (or e) can only mark post-verbal noun phrases. Thus, clauses involving the sequence VS follow an ergative-absolutive case-marking pattern, where ergativity is, as in many Polynesian languages, a fairly "shallow" phenomenon (Besnier 1981a), while SVO clauses follow a nominative-absolutive pattern, in which case relations are retrievable from the order of the constituents.

Table 1 below summarises the word-order configurations attested in Tuvaluan, and the case-marking strategy associated with each variant.

Table 1: Attested and unattested word-order configurations

intransitive clause	O/1-NP transitive clause	2-NP clause
VS SV VeS[4] V	VO OV VeS V	VeSO VOeS OVeS SVO
	*SV	*SOV *OSV

(e = ergatively marked; * = unattested)

The Tuvaluan dialects, unlike other Polynesian languages of the area, do not make productive use of pre-verbal clitic pronouns (although they are more productive in some dialects than in others); their use is restricted to a small number of set expressions, which are not relevant to the present discussion (see Besnier 1986 for further details). Furthermore, the suffixation of verbs with -gina (or, alternatively, -gia in Southern Tuvaluan), the productive reflex of the Proto-Polynesian suffix *-(C)ia (Clark 1976, 1977, Chung 1978, etc.), which, in Tuvaluan, "boosts" the transitivity of the verb, does not appear to have any effect on word-order configuration.[5]

It is important to distinguish between unmarked clause-initial noun phrases and another type of clause-initial noun phrase, marked with ko, the latter type being much more common in Polynesian languages than the former. The morpheme ko, whose exact role appears to vary from one language to the other, marks, amongst other things (such as predication), new-information focus in Tuvaluan, whether contrastive or not, as illustrated in (3). The term "focus" is used here to refer to "the essential piece of new information that is carried by a sentence" (Comrie 1981:57).

(3) STV E isi ttino Nukufetau, e fakaigoa kiaa Taukiei. I suaa
Nps *exist the+man Nukufetau* Nps *named after Taukiei at other*

taimi laa, ko tena tama e nofo i Nukulaelae nei, fai mai ei kee fanatu.
time then Foc *his child* Nps *stay on Nukulaelae this say* Dxs Anp Sbj *go*+Dxs

There was a Nukufetau man named Taukiei. On one occasion, he told his son, who was living here on Nukulaelae, to come over.

The association of ko with the marking of new information is illustrated by the fact that ko-marked structures are typically encountered in answers to

information questions (a prototypical context for new information), in which the use of an unmarked preposed noun phrase, in contrast, would be infelicitous. As will be shown later, the pragmatic role associated with ko-marked noun phrases is thus diametrically opposed to that associated with unmarked noun phrases. Furthermore, ko-focused structures are, syntactically speaking, clefted constructions, while unmarked nominal arguments are not.

2. THE SYNTAX OF WORD ORDER

Despite the relative surface word-order freedom, evidence exists for positing VSO as both the underlying order and the syntactically basic order. The argument that follows echoes in part a similar discussion for Kapingamarangi by Chung (1978:15-20).

Excluding the possibility that Tuvaluan has the typologically marked OVS and VOS orders as underlying word order,[6] and excluding verb-final structures because of the distributional restrictions noted above, SVO and VSO are the only possible underlying word orders for Tuvaluan syntax. It is to be noted that the implicational universals proposed by Greenberg (1966), Vennemann (1974 and 1975), and Hawkins (1983) do not provide any evidence for treating either order as basic. As pointed out by Hawkins (1983:16), word-order universals only contrast verb-initial and verb-final structures, leaving verb-medial systems to follow the patterns of one or the other.

First of all, in order to account for the relatively frequent object-initial and subject-final word-order variants (i.e. OVS), one needs to posit two separate movement processes if SVO is posited as the underlying word order (i.e. a rule of subject postposing and a rule of object preposing), while only one such process is needed if VSO is the underlying shape of the clause. The simplicity of the latter account provides one argument for preferring a verb-initial order to a verb-medial one.

Furthermore, subordinate clauses do not allow any nominal constituent to precede the verb. In such clauses, the order of the sentential constituents is strictly verb initial, whether the subject precedes the object or not. Compare, for example, the grammaticality of (4a), where the subordinate clause is verb initial, with its ungrammatical equivalent in (4b), in which the subject of the subordinate is preposed to the verb:

(4) STV

a. Toku tamana ne fai mai kee fanatu koe. VS
my father Pst *say* Dxs Sbj *go*+Dxs *you*
My father said that you should come over.

b.*Toku tamana ne fai mai koe kee fanatu. *SV

This is reminiscent of a similar constraint on subordinate clauses in German and Dutch, in which these follow a strict verb-final order. This fact has been argued by Bierwisch (1963) for German and Koster (1975) for Dutch to be a strong argument for positing the underlying syntax of these two languages as being of the verb-final type.

Relativisation provides further evidence for the verb-initial nature of the underlying syntax of Tuvaluan. One of the two relativisation strategies found in Tuvaluan grammar consists in replacing the relativised noun phrase by a

resumptive pronoun, as illustrated in (5a) below. The resumptive pronoun preferably precedes the direct object of the relative clause (VOS relative clauses, as in (5b), are unidiomatic); however, it cannot precede the verb of the relative clause, as shown by the ungrammaticality of (5c), even though the boundary between the relative clause and the head may be marked with an optional relativising demonstrative teelaa (singular relative head) or kolaa (non-singular relative head):

(5) NKL

a. Teenei ttino e hai saale nee ia aku fekau. VeSO
this the+man Nps *do often* Erg *he my errand*
This is the man that does my errands.

b. ?Teenei ttino e hai saale aku fekau nee ia. ?VOeS

c. *Teenei ttino teelaa a ia e hai saale (nee ia) aku fekau. *SV(eS)O
this the+man that he Nps *do often* Erg *he my errand*

The resumptive-pronoun relativisation strategy thus suggests that relativisation applies to a verb-initial structure, and not to a verb-medial configuration.

Main or independent transitive clauses in which the subject is preposed to the verb also involve pronominal copies. In these, the transitive subject is optionally co-referenced with a post-verbal ergatively-marked pronoun, as illustrated in (6). This type of construction is not only possible, but very frequent in unelicited discourse, and, in elicited sentences, judged to be preferred to corresponding structures without a post-verbal pronominal copy.

(6) NTV Te loomatua ni fakavela e ia te tii. SVeSO
the old-woman Pst *heat-up* Erg *she the tea*
The old woman heated the tea up.

It thus appears, from the distributional patterns exhibited by these resumptive pronouns, that preposed transitive subjects are the result of a movement rule that leaves an optional pronominal trace in the underlying post-verbal position.

Similarly, the shape of imperative clauses in Tuvaluan suggests that imperative-formation is a process that discriminates between subject-initial and verb-initial structures. Imperative clauses are formed by deleting the tense-aspect marker immediately preceding the verb. The subject of an imperative clause need not be affected by imperative-formation as long as it follows the verb, as illustrated in (7a); in contrast, no pre-verbal subject is allowed to surface in imperative clauses, as attested by the ungrammaticality of (7b):

(7) STV

a. Olo koutou keaattea ia maatou e fakalavelave! VS
go you-3 *away because we*-3 Nps *busy*
Go away, we are busy!

b. *Koutou olo keaattea ia maatou e fakalavelave! *SV

Imperative formation, thus, provides further motivation for treating the underlying word order of Tuvaluan syntax as verb initial.

To summarise, the four features of Tuvaluan syntax presented above all converge to yield the same conclusion: that the syntax of the Tuvaluan dialects is underlyingly verb initial.

Furthermore, there exists strong evidence for treating VSO as the least marked, or most basic constituent order. Typologists like Hawkins (1983) distinguish the notion of underlying word order from that of basic word order, the former being only one of the several criteria involved in the determination of the latter. According to Hawkins, three sets of criteria can be used to define basicness: overall frequency, structural frequency (i.e. the relative range of occurrence of a word-order variant over the different types of syntactic structures), and grammatical markedness (1983:12-16). As illustrated above, VSO is the structurally most frequent variant; it is also treated by the syntax (through the pronoun-resumption processes, for example) as the least marked variant. The fact that VSO satisfies these two criteria is sufficient motivation for treating it as the basic word order, despite the fact that the last criterion, overall frequency, is only satisfied for the more formal types of discourse.[7]

3. THE PRAGMATICS OF WORD ORDER

What, then, are the factors governing the occurrence of the various word-order configurations? In the light of Li and Thompson's (1976) typological criteria for distinguishing subjects from topics, the clause-initial nominal constituent in Tuvaluan presents itself as a prototypical old-information topic slot. Tuvaluan, for instance, exhibits "double subject" constructions (Li and Thompson 1976:480-481), in which an intransitive sentence surfaces with two nominal arguments. In such sentences, which are frequently encountered in languages in which topics play an important role, whether statistically or grammatically, the first noun phrase is a topic, syntactically independent of the rest of the clause. In Tuvaluan, these structures are most commonly of the shape NP V Poss N, where the first noun phrase and the possessive marker are co-referential:

(8) NEA Aku koa mmae toku piho. NP-VS
I Inc *hurt my head*
I have a headache.

(9) STV A ttamaa toeitiiti koo too tena moe. NP-VS
Cnt *the+man*+Spc *almost* Inc *fall his sleep*
That guy almost fell asleep.

Li and Thompson further propose a number of cross-linguistic characteristics of topics. First, topics are often obligatorily definite; the clause-initial position in Tuvaluan is indeed restricted to definite noun phrases, as illustrated in (10) and (11), both of which are ungrammatical because the pre-verbal noun phrase is indefinite:

(10) NTV *Ni falaoa ni kai e aku. *OVeS
some bread Pst *eat* Erg *I*
(I ate some bread.)

(11) NTO *Se tino ni kai (e ia) te ika teelaa. *SV(eS)O
a person Pst *eat* Erg *he the fish that*
(Someone ate that fish.)

As further evidence for this restriction, we note that the subject of a possessive clause (i.e. the noun phrase referring to the possessed participant) is always indefinite when postposed to the verb; when moved to the beginning of

the clause, however, it is always marked for definiteness. Compare the following contrastive sentences, the second of which was uttered by someone who knew that the addressee had a hatchet (and hence marked the nominal element for old-information topicality). The possessive construction in (12b) is used as an indirect request:

(12) STV

a. E isi sou takuu? VS
Nps *exist your hatchet*
Do you have a hatchet?

b. Tou takuu e isi? SV
your hatchet Nps *exist*
Is your hatchet handy?

Li and Thompson's second criterion states that a topic need not have any selectional relation with the verb of the clause. The "double subject" constructions illustrated above show that this criterion applies to at least one type of topicalised structure in Tuvaluan.

Criterion 3 states that, cross-linguistically, topic selection does not depend on the verb. As illustrated by many of the above examples, Tuvaluan verbs play no role in determining which noun phrase in the clause gets fronted to topic position, this being determined entirely by the functional role played by the noun phrase in the discourse.

Verb argeement, also, is not triggered by the topic. In Tuvaluan, when the verb agrees in number with a nominal constituent (agreement being marked by consonant gemination), it is with the (intransitive) subject, and not with the clause-initial noun phrase; this is Li and Thompson's criterion 5. In sentences (13a-d), whether the verb is to be interpreted as transitive or intransitive is determined by whether agreement with the overt noun phrase takes place or not; the position of the nominal constituent in the clause has no effect on agreement:

(13) STV

a. Ana puaka koo oti ne kkai katoa. (agreement) SV
his pigs Prf *eat all*
His pigs have all eaten.

b. Koo oti ne kkai katoa ana puaka. (agreement) VS

c. Ana puaka koo oti ne kai katoa. (no agreement) SV
his pigs Prf *eat all*
His pigs have all been eaten.

d. Koo oti ne kai katoa ana puaka. (no agreement) VS

Finally, grammatical processes in Tuvaluan syntax refer, not to clause-initial constituents as a grammatical category, but to the categories defined by grammatical relations and, to a lesser extent, surface case (Besnier 1981a). One illustration will be provided here: quantifier float can be triggered from non-oblique nominal constituents (intransitive subjects, transitive subjects, direct objects), irrespective of their position in the sentence, as illustrated in (14a-c); in contrast, a quantifier cannot be launched from an oblique noun phrase, irrespective again of its position in the sentence ((14d-f)):

(14) STV

a. Ne ttogi aku naifi katoa i te sitoa. V[Oq]
Pst *buy my knife all at the trade-store*
I bought all my knives at the trade store.

b. Ne ttogi katoa aku naifi i te sitoa. [Vq]O

c. Aku naifi ne ttogi katoa i te sitoa. O[Vq]

d. I aso katoa e fano o fai uttanu. [NPq]V
on day all Nps *go* Cmp *do sprouted-coconut*
Every day he goes to gather sprouted coconuts.

e.*I aso e fano katoa o fai uttanu. *NP[Vq]

f.*E fano katoa i aso o fai uttanu. *[Vq]NP

All processes of Tuvaluan syntax behave in a fashion similar to quantifier float with respect to the position of the nominal constituents in the clause. This satisfies Li and Thompson's criterion 7.

The above discussion shows that, in Tuvaluan, all clause-initial nominal constituents are topics (the same set of arguments can be shown to also apply to preposable oblique non-phrases), that the high incidence of non-basic configurations is due to the salience of the notion "topic" in Tuvaluan discourse, and, thus, that word-order freedom is pragmatically governed.

4. EXPLAINING *SOV, *OSV, AND *SV

It is now possible to provide a functional explanation for the absence of SOV, OSV, and transitive SV structures from the range of word-order possibilities. We shall first turn to the most readily accountable of these, i.e. the ungrammaticality of SOV and OSV configurations.

First of all, given the fact that the order of non-sentential constituents in Tuvaluan (noun-adjective, preposition-noun, etc.) is characteristic of a verb-initial grammar, the verb-final variants SOV and OSV, if they existed, would violate Hawkins' (1983) "principle of Cross-Category Harmony": according to this principle (and any of the variants that have been proposed for it - Lehmann 1973, Vennemann 1975, etc.), a language with basic verb-initial syntax will be typologically most "consistent" (Smith 1981) if it also has post-nominal adjectives, post-nominal genitives, and so on. Thus, in the paradigm of possible word-order configurations, SOV and OSV are the most anomalous from a typological perspective.[8]

Furthermore, if SOV and OSV were attested, they would be the result of the topicalisation of both the transitive subject and the direct object of the clause. Tuvaluan grammar, however, strongly constrains the occurrence of double topics in a clause: these may only occur if the first noun phrase is separated from the rest of the sentence by an intonational break, an adverbial determiner, and a demonstrative, and if that first noun phrase is a highly oblique, non-obligatory term of the sentence, such as, for example, a temporal or locative noun phrase (the second noun phrase, on the other hand, must be a non-oblique noun phrase):

(15) NKL Te aso teelaa eiloa, a Toe ne vau kia aku. NP-SV
the day that indeed Cnt *Toe* Pst *come to me*
The other day, Toe came up to me.

The status of the first noun phrase in these constructions is that of left-dislocated elements (Duranti and Ochs 1979), a syntactic position to which only certain types of oblique noun phrases have access (such as locatives, temporal expressions, and some instrumentals).

These restrictions on the co-occurrence of more than one pre-verbal nominal constituent thus explain the fact that constructions with both a pre-verbal transitive subject and a pre-verbal direct object are not attested: a clause with two pre-verbal noun phrases requires one of these to be left-dislocated, a position to which non-oblique noun phrases do not have access.

Transitive SV constructions, if they were attested, would be the result of two separate processes: the preposing of the transitive subject to topic position, and the zero-pronominalisation (Besnier 1985) of the direct object. In terms of the "Hierarchy of Informational Value" proposed by Lakoff (1968) and refined by Cole (1974), a zero pronoun is the least explicit mode of reference for a nominal argument, and, thus, the most likely to refer to old information. However, it is well documented that, cross-linguistically, subjects are more likely topics than objects, all other things being equal (Givón 1977, Keenan 1976). Tuvaluan grammar may thus regard as pragmatically unnatural and grammatically anomalous a configuration that places an already highly topical argument (the subject) in a position where its topicality is further stressed, but where the object is marked for even greater topicality than the subject.[9]

5. WORD ORDER AND CASE MARKING

Other facts about Tuvaluan syntax, however, suggest that this functional explanation is not sufficient, and that other factors relating to the nature of case marking are at play. It will be shown here that the contrast between preposed transitive subjects and postposed transitive subjects involves not only the pragmatic notion of topicality, but also some of the semantic notions associated with high-transitive subjecthood. This conclusion is suggested by the co-occurrence of certain syntactic constraints involving word order with the semantic characteristics of these structural types, which will first be presented in this section.

As mentioned in section 1 above, no nominal constituent in the Tuvaluan clause is marked for case other than post-verbal transitive subjects (and a few post-verbal intransitive subjects, as will be shown below), which are marked with the ergative preposition nee (e in Northern Tuvaluan). The fact that this preposition can mark appropriate subjects only if they are post-verbal is important for the word-order syntax of Tuvaluan, as I shall demonstrate here.

In Tuvaluan, there is a class of verbs that can be used in the main clause of complex sentences in which the subject of the subordinate clause may be raised to the main clause (arguments for treating the resulting clauses as being derived through raising will not be presented here). With a small subclass of these main-clause verbs, the raised subject is obligatorily marked for the ergative case, whether the subordinate clause of which it is the underlying subject is transitive or not; these verbs behave syntactically like the Niuean verbs labelled "TS-ergative" by Levin and Massam (1986).[10] The verb maua *able to* is one such verb; in (16), the raised subject of the intransitive compound ("object-incorporated") verb ssali kaleve is marked for the ergative case, this case assignment being required by maua in the upper clause:[11]

(16) STV E maua nee ia o ssali kaleve. [VeS[V]]
Nps *able* Erg *he* Cmp *draw toddy*
He knows how to tap coconut toddy.

This subclass of raising verbs (which will be referred to as maua-type verbs) has two other important properties. First, the subclass includes the only raising verbs that denote notions involving the high agentivity of the subject in the action or state of the verb (capability, knowledge, and memory - see note 11 for an exhaustive list), where "high agentivity" is a cover term for volition and/or a high degree of direct physical activity and/or the affectedness of the object by the agent (i.e. the subset of transitivity parameters proposed by Hopper and Thompson 1980 that refer to the agentive nominal constituents). Secondly, raised subjects of maua-type verbs are rarely preposed to topic position (a position in which they would lose their morphologically overt case marker).

The second characteristic is not a restriction on the topicalisation of raised subjects. Indeed, raising verbs outside of this subclass (i.e. raising verbs that do not mark the raised subject ergatively) do not impose any restriction on the topicalisation of the raised subject. Furthermore, the restriction on the topicalisation of raised subjects is not an absolute rule with maua-type, but, rather, a tendency: in certain circumstances, such as when the raised subject is a first- or second-person singular pronoun (hence highly topical), it may be preposed, in which case it is interpreted as emphatically contrastive, as illustrated in (17):

(17) STV Koe naa e maua o kake? [SV[V]]
you that Nps *can* Cmp *climb*
How about you, can YOU climb (trees)?

The properties exhibited by maua-type verbs suggest that these verbs tend to be accompanied by postposed (and, hence, ergatively-marked) subjects. This fact is posited to be associated with the semantics of these verbs, or, more precisely, with the fact that the actions or states they denote *require* that their grammatical subject be marked for high agentivity, a requirement that is fulfilled by the ergative marker nee.

Yet another type of structure, which will be called pseudo-ergative structures, suggests the same analysis. Pseudo-ergative structures involve an intransitive verb whose post-verbal subject is marked ergatively, as illustrated in (18) and (19):

(18) STV A papa koo manogi ssogo ne mimi nee te puusi. [SV][VeS]
Cnt *mats* Inc *smell-of-urine* Pst *urinate* Erg *the cat*
The mats smell bad because the cat has been urinating all over them.

(19) NGA Taatou koa hihiu ni too e te vaiua. [SV][VeS]
we-3 Inc *wet* Pst *fall* Erg *the rain*
We were going to be rained on and drenched.

In these examples, the verbs mimi and too are canonically intransitive verbs, as they are not used transitively elsewhere. Transitivity is a well-defined concept in Tuvaluan syntax, in that a number of tests can be devised to distinguish highly transitive constructions from constructions that are low in transitivity (Besnier 1981a). The intransitivity of pseudo-ergative structures

is illustrated, for example, by the fact that verbs suffixed with the detransitivising suffixes of shape -(C)ia may be found in this type of construction:

(20) STV A ia koo onosia nee te fenua. OV-CiaeS
Cnt *he* Inc *ostracised* Erg *the island*
He is ostracised by everyone on the island.

(21) NTO Aku ni osofia e toku taina. OV-CiaeS
I Pst *pounced-on* Erg *my brother*
I was pounced on by my brother.

Furthermore, the fact that these constructions are "frozen" in an (NP)VeS configuration (where NP is the "affected" noun phrase, and S the ergatively-marked noun phrase) indicates that they are not transitive constructions, since the latter type always have SV(eS)O, VeSO, and VOeS alternatives. Further evidence for treating these constructions as intransitive constructions is provided by the fact that they do not undergo the nominalisation patterns associated with transitive clauses, and that their ergatively-marked participant cannot launch a quantifier, a process that all other transitive subjects can trigger (Besnier 1981a). Unlike canonical intransitive clauses, however, pseudo-ergative structures do not undergo subject-verb agreement, and, thus, do not behave entirely like intransitive constructions.

At the surface (case-assigning) level of syntax, the ergative case in pseudo-ergative structures thus denotes, not transitive subjecthood (the grammatical relation usually associated with ergativity), but high agentivity. Indeed, pseudo-ergative structures always denote situations in which the role of the ergatively-marked noun phrase is one of high, often negative, affect on another entity of the discourse. The association of a case-marking pattern with a semantic concept other than the usual notions of subject or object is reminiscent of the "split intransitive" (Merlan 1985) or "active" (Harris 1982) case-marking systems found in certain Amerindian languages, in which the subject of an intransitive verb is marked like a direct object if non-volitional, or like an intransitive subject if volitional.[12]

Returning to canonically transitive constructions, similar types of semantic contrasts can be established for clauses in which there is a choice of case-marking strategy. In pragmatically unmarked independent clauses, that is, in clauses where the nominal arguments are not in any particularly prominent topical position, the choice between a preposed, neutrally-marked transitive agent, and a postposed, ergatively-marked transitive agent is not only determined by topicality, as established in section 3, but also by the degree of agentivity of the subject. Thus, the high-low agentivity contrast is another of the determining factors distinguishing between preposed (morphologically unmarked) transitive subjects and postposed (ergatively-marked) transitive subjects, as attested by the contrast between (22a) and (22b) below:

(22) STV
a. Ne maua nee au a ika konei annafi. VeSO
Pst *get* Erg *I* Cnt *fish these yesterday*
I caught these fish yesterday (with my own hands).

b. Au ne maua a ika konei annafi. SVO
I Pst *get* Cnt *fish these yesterday*
I obtained these fish yesterday (from someone, etc.).

Thus, the ergative case marker used in pseudo-ergative constructions and with maua-type verbs is not simply homophonous with the ergative marker used in independent clauses; rather, it marks, in all these syntactic contexts, the same type of semantic relation of the subject to the verb.

Furthermore, since topichood and high agentivity are not mutually exclusive concepts, the same subject may appear on both sides of the verb, as in sentence (22c), a commonly attested structure as pointed out earlier:

(22) STV
c. Au ne maua nee au a ika konei annafi. SVeSO
I Pst *get* Erg *I* Cnt *fish these yesterday*
I was the one who caught these fish yesterday.

Finally, to return to the question of the ungrammaticality of transitive SV structures, Tuvaluan discourse exhibits a strong tendency to reduce transitive clauses to structures consisting of a verb and a single noun phrase. In fact, two-argument transitive clauses are very infrequent in natural discourse.[13] Several strategies are used to achieve this preferential clause structure: one of the nominal arguments in transitive structures may be zero-pronominalised (as discussed earlier); the transitive agent may be referenced within the object phrase by a possessive pronoun, as in example (23) below;[14] or a complex structure can be used, as in example (24), in which the subject is referenced overtly in the main clause, while only the object is overt in the subordinate clause:

(23) STV Taku ika teenei ne ffuti annafi. OV
my fish this Pst *pull yesterday*
I caught this fish yesterday.

(24) NEA A Ioane ni haga o ffuti te ika teelaa. [SV[VO]]
Cnt *Ioane* Pst *apply-oneself* Cmp *pull the fish that*
Ioane (applied himself and) caught that fish.

From a semantic point of view, the difference in the meaning of these "alternative" clause types and their two-argument, main clause paraphrases is minimal.

An important consequence of these tendencies is the fact that there is a choice, in any transitive sentence, as to which of the two non-oblique nominal constituents is to be expressed overtly. If a transitive subject is the overtly expressed nominal constituent, there always is a strong tendency to interpret its overt realisation as denoting high agentivity, as in example (25):

(25) STV Teenaa laa, koo oti ne vvae nee ia. VeS
thus Prf *divide* Erg *she*
So SHE has already divided (it).

Indeed, if the transitive subject is to be expressed overtly but not marked for high agentivity, it is normally expressed through zero-pronominalisation, through a possessive marker, or through any other way that would not require it to be marked ergatively. Thus, a transitive subject serving as the sole constituent of a clause must be postposed to the verb, where it is marked for the ergative case to denote high agentivity.

6. IMPLICATIONS AND CONCLUSION

It has been shown that word order in Tuvaluan is used to mark the pragmatic structure of the clause, with the clause-initial (pre-verbal) nominal element denoting topicality. Furthermore, subjects marked for the ergative case (because they are raised to a main clause with a maua-type verb, or because they are within a pseudo-ergative construction, or because they are post-verbal transitive subjects) denote high agentivity, in contrast to the various other ways of denoting subjecthood. Word-order variation in Tuvaluan, thus, requires a complex explanation, involving the pragmatic structure of the sentence on the one hand, and the semantic value of the ergative morpheme on the other hand.

It is commonly assumed in the literature on word-order change that word-order variability, particularly when it is as widespread as in Tuvaluan, is inherently unstable, and, in the unmarked case, leads to the reanalysis of one of the word-order variants as the new basic order of the syntax. This assumption is evident in the various motivations that have been proposed as triggering word-order changes: the resolution of relational ambiguity at the sentence level (Lehmann 1973, Vennemann 1975), the reanalysis of topichood as subjecthood (Givón 1977), the reanalysis of "afterthoughts" as objects (Hyman 1975), and the competition of different word-order configurations between main and subordinate clauses (Parker 1980) all involve, in one way or another, the notion that grammatical optionality, which is more often than not created by the encoding of pragmatic information in the sentence, is a transitional state.[15] To use Givón's (1979) words, yesterday's pragmatics is today's syntax.

Is Tuvaluan undergoing a change from VSO to SVO involving the "grammaticalisation" of its pragmatic structure? I suggest here that, as far as one can argue for or against change in process, this is not the case.

Positing such a change in process would indeed be a tempting analysis. From a typological perspective, historical evolution from VSO to SVO is both common and natural: word-order changes involving the reanalysis of the topic slot as the subject slot have been reported to have occurred in Indo-European languages (Vennemann 1974) and Semitic languages (Givón 1977); and the change from verb-initiality to verb-mediality is, according to Keenan's (1979) "Subject-Frontness Hierarchy", given in (26) below, part of a wider change motivated by both processing factors and typological frequency:

(26)

[less "preferred"]
------->

SVOX > VSOX > VOSX > VOXS

<-------
[pressure to change]
[typological frequency]

Keenan's (1979) Subject-Frontness Hierarchy

However, Tuvaluan does not appear to fit any of the descriptions of transitional systems that have been proposed to date. Furthermore, the complex interaction of semantic and pragmatic factors in word-order choice appears to indicate that the system is in fact stable.

First of all, we note from Table 1 that the Tuvaluan word-order system is such that the only configurations that would involve relational ambiguity are ungrammatical. The identification of the subject and the object depends on the

following principles: if a transitive verb has only one morphologically unmarked nominal constituent (whether it be preposed or postposed), it is an object; if two morphologically unmarked nominal constituents are present (in a NP-V-NP configuration), the first is to be interpreted as the subject, the second as the object; finally, in all other possible cases, the subject is marked with nee. Thus, the Tuvaluan system exhibits no possible relational ambiguity that would motivate a word-order change.[16]

Secondly, it was shown in this paper that word order is dependent both on the relative topicality of each nominal constituent, and on whether the agent is to be marked for high agentivity. The high topicality of any nominal constituent is marked by preposing it to the sentence-initial position, while high agentivity is associated with the post-verbal position. Thus, if the topic slot is to be reinterpreted as a subject slot, the grammar will have to forgo its current capability of marking a subject for high agentivity, since subjects will be obligatorily pre-verbal. Object topicality will also have to be marked in some way other than through a movement process, and it will have to reanalyse the currently ungrammatical transitive SV structures as grammatical. This topic-as-subject reanalysis, thus, is structurally very "costly", and is unlikely to be taking place.

Finally, no morphologically complex structure is involved in the word-order inventory of Tuvaluan (such as the verb serialisation in Chinese, posited by Li and Thompson as the trigger for word-order change in that language). Nor is there a configurational discrepancy between the basic and underlying word order of main and subordinate clauses: the difference between main and subordinate clauses is that, while word-order variations are possible in the former, they are not in the latter, a state of affairs that appears to be common in Australian languages (Mallinson and Blake 1981:129).

The Tuvaluan word-order system thus appears to be a highly "efficient" system, in that it allows the encoding of an appreciable amount of semantic and pragmatic information with a minimal amount of morphology, while avoiding any possibility for grammatical ambiguity. Furthermore, motivational elements traditionally associated with diachronic processes of word-order change are not attested in Tuvaluan. The system therefore is a diachronically stable phenomenon, in contrast to the situations of word-order change in process documented in the literature. In Tuvaluan, yesterday's pragmatics remains today's pragmatics.

To conclude, it is suggested here that many of the same patterns presented here for Tuvaluan may also be characteristic of at least some of the Polynesian Outlier languages. Indeed, the same degree of word-order variability is found to be at play in several of these languages. This suggests that, in those languages, word order also marks pragmatic structure. Consider the following examples, from Anutan, Tikopian, Takuu, Luangiua, and West Futuna-Aniwa respectively, in which the same degree of variability in word order is illustrated, and in which the same patterns of interaction between case marking and word order appear to be at play (at least in the languages with ergative case-marking):

(27) ANU (Feinberg n.d.)

a. Te penua ne oro o taa te marara. [SV[VO]]
the people Pst *go* Cmp *strike the charcoal*
The people went and painted themselves with charcoal.

b. Nga manumanu ne taamate e Motikitiki. OVeS
the animal Pst *kill* Erg *Motikitiki*
Motikitiki killed the animals.

c. Natou umu ne tao. OV
their oven Pst *bake*
(They) baked (the food in) their underground oven.

(28) TIK (Early 1981:114-118)

a. Te ua ne too, te raa ne saa. SV
the rain Pst *fall the sun* Pst *shine*
The rain fell, the sun shone.

b. A taagata ne kai te ufi. SVO
the men Pst *eat the yam*
The men ate the yam.

c. Te ufi, ne sori e a kuou ki ei. OVeS
the yam Pst *give* Erg Art *I to* Anp
I gave him the yam.

(29) TAK (Reedy 1977)

a. Te lani raa e uri. SV
the sky that Nps *blue*
The sky is blue.

b. Te poi raa e oso te manu. SVO
the dog that Nps *catch the bird*
The dog caught the bird.

c. Te poi raa e osofia te manu. SV-CiaO
the dog that Nps *catch*+Cia *the bird*
The dog caught the bird.

d. Te manu raa e osofia te aa? OV-CiaS
the bird that Nps *catch*+Cia *the what?*
What caught the bird?

(30) LUA (Salmond 1974)

a. Ke poi la me a'e ke uŋa la. SV
the dog that say up the crab that
The dog said (to) the hermit crab ....

b. La me a'e ke uŋa la. VS
then say up the crab that
Then the hermit crab said ....

c. Keeŋaa ko'o lua la e 'ila ka'upu hoo kahi. SVO
thus Num *two those* Nps *look girl* Num *one*
The two looked intently at one of the girls.

d. Sipuŋi e moe se loŋo ŋaa kamali'i kaahao vaa loko ke manava
Sipuŋi Nps *sleep* Neg *feel the children play in inside the belly*
aia la. [SV][VO]
his that
Sipuŋi slept, not feeling the children playing inside his belly.

e. Keeŋa ŋoo loŋo ka'upu la. [VS(O)]
then Cmp *feel girl that*
Then the girl felt (something).

(31) WFU (Dougherty 1983)

a. Ko to te ua. VS
Inc *fall the rain*
It's raining.

b. Ta kiri pepa ni safifi. SV
the skin paper Pst *curl*
The sheet of paper has curled.

c. Ta tao nei kofia te kamkama. SVO
the spear Pst *pin the crab*
The spear pinned the crab.

d. A tama Pau ne fujia e kirea ta uorukago. SVeSO
the people Pau Pst *catch* Erg *they the tuna*
The people of Pau caught, they did, a tuna.

e. Ta fakau ro fano, avau kan fakaoa. [O][SV]
the person Irr *go I* Fut *reward*
The person who goes, I will reward (him).

Dougherty (1983:119) analyses the double mention of the agent in example (31d) as correlating "with the emphasis on the potency of the subject", which is the same claim as that made here for the corresponding phenomenon in Tuvaluan. West Futunan appears to differ from Tuvaluan, however, in allowing the clefting of an object noun phrase immediately before a preposed subject (sentence (31e)).

In contrast, for a number of Polynesian Outlier languages, the only attested word-order configuration is SVO. In those languages, verb-mediality appears to have become the only possible word-order configuration, and, presumably, the syntactically basic order. This is illustrated by Mele, an Outlier spoken in the Vila Bay of Efate Island, Vanuatu:

(32) MEF (Clark 1975a and b)

a. Napoonaa ti tamaffine naa kuu-tere. SV
then the girl that Inc *run*
Then the girl started running.

b. T'nuufine takua kaia waawa neana. SV
The+old-woman say to uncle her
The old woman told her uncle.

c. Maasai raaraku te paki. SVO
Maasai unload the canoe
Maasai unloaded the canoe.

Crucially, the Outlier languages that are documented as having reanalysed their basic word order to strict verb-mediality are precisely those with a long history of contact with non-Polynesian Oceanic ("Melanesian") languages which have a strict SVO surface syntax: Mele-Fila, Emae, and perhaps some other Outliers. The influence of non-Polynesian languages on these Outliers, as shown by Clark (1978, 1986), is very strong at all levels of structure. Interestingly, West Uvea, despite the fact that it has received considerable influence from Iaai, a neighbouring non-Polynesian language, does not exhibit the strict SVO ordering attested for Outliers of the Mele type, as illustrated in (33) below, a feature that can probably be explained by the fact that Iaai is essentially VOS (Moyse-Faurie and Ozanne-Rivierre 1983):

(33) WUV (Besnier, field notes)

a. E kitea a de kulii de hoto. VeSO
Nps *see*+Trn Erg *the dog the hen*
The dog sees the hen.

b. De kulii e/ide kitea de hoto. SVO
the dog Nps/*he*+Nps *see*+Trn *the hen*
The dog sees the hen.

c. De hoto e kitea a de kulii. OVeS
d. De hoto e kitea. OV

In the light of Smith's (1981) remarks to the effect that word order is a syntactic feature that is very readily diffused across genetic boundaries, we may thus infer that the strict verb-medial systems developed by languages of the Mele type is the result of non-Polynesian influence, and, thus, is independent of the pragmatically-governed word-order variability exhibited by Tuvaluan, Tokelauan (Hooper 1986), and some Outliers.[17]

NOTES

1. This paper is based on field work conducted in Tuvalu in 1980-82. I thank Pat Clancy, Bernard Comrie, Ed Finegan, Jack Hawkins, Will Leben, and Elizabeth Traugott for having greatly contributed, at various stages, to the development of the ideas presented here. This paper also benefited from suggestions and comments by James Alexander, Joseph Finney, Michael Goldsmith, Ray Harlow, Robin Hooper, Ken Cook, Jacob Love and Françoise Rivierre, and by various members of the Linguistics Department at the University of Southern California, where it was also presented in September 1984. I am grateful to the Fondation de la Vocation (Paris) for supporting part of the field work on which this paper is based and to the Government of Tuvalu for permission to conduct research.

2. The Polynesian Outlier languages are spoken on a set of widely dispersed islands of Melanesia and Micronesia, namely: Takuu, Nukumanu, Nukuria (in Papua New Guinea's Northern Solomon Province), Luangiua, Sikaiana, Rennell, Bellona, Tikopia, Anuta, Pileni-Taumako (in the Solomon Islands), Mele-Fila, West Futuna-Aniwa, Emae (in Vanuatu), West Uvea (in New Caledonia's Loyalty Islands), and Nukuoro and Kapingamarangi (in the Federated States of Micronesia).

3. Examples are preceded with a three-letter abbreviation of the name of the dialect from which they are taken, a list of which may be found in the appendix, along with a key to the abbreviations used in interlinear glosses. Most contrastive sets of examples cited in this paper were elicited; non-contrastive examples were all taken from spoken or written textual sources.

4. This structural type is that of "pseudo-ergative" constructions, discussed in section 5.

5. In a survey of word-order variation in the Nanumaga dialect of Tuvaluan, Finney (1983) focuses precisely on one of the few verbs (i|oa *to know*) that allow clitic subject pronouns. The number of grammatical and ungrammatical variants for this type of verb is thus much greater than those considered in this paper (in that one has to allow for the co-occurrence of a clitic pronoun with a full noun phrase, with a free pronoun, etc.).

6. Two languages spoken in the immediate neighbourhood of the Tuvalu, Gilbertese (Kiribati) and Standard Fijian, are often quoted in the literature as instances of VOS languages. Little data on Gilbertese syntax is currently available, although Cowell (1951) and Trussel (1979) both describe Kiribati transitive clauses as strictly following a VOXS order. It is much less clear that Fijian syntax is that of a VOS language than it is usually assumed. Some of the problems associated with positing Fijian as a subject-final language are pointed out by Keenan (1978). Geraghty (1983) suggests that the statistical predominance and unmarked nature of VOS is not a characteristic of any dialect of Fijian, but of Pidgin Fijian, of foreigner talk and of the Fijian obtained in linguistic elicitation situations (see also Geraghty 1978 and Moag 1978). The VOS hypothesis for Fijian is, according to Geraghty, "a fabrication wrought by generations of informants and grammarians guided more by translations than by spoken Fijian" (1983: 391). Further research is needed on the subject. It is a fact, however, that VOS is reported as the basic order of many languages of the greater Austronesian area, for example: Palauan (Georgopoulos 1986), Toba Batak (Cumming 1986), Malagasy (Keenan 1978), and many New Caledonia languages (Moyse-Faurie and Ozanne-Rivierre 1983).

7. This will not be illustrated here. A preliminary statistical investigation of the problem is outlined in Besnier 1981a; a more sophisticated computer-assisted stylistic analysis of word-order variation is in progress.

8. Verb-final variants are nevertheless attested in languages that are basically verb initial, as illustrated by the actor-emphatic construction in Maori and other Eastern Polynesian languages (Harlow 1986).

9. The Government and Binding framework developed by Chomsky (1981) and refined by his students would provide the following syntactic explanation for the non-occurrence of transitive SV structures in Tuvaluan: it has been shown by Huang (1983) that, in zero-pronominalising languages like Chinese, zero objects are not "true" pronouns, but traces (empty categories) bound by non-realised (deleted) object topics; thus, transitive SV clauses would involve the co-occurrence of an overt subject topic and of a non-overt object topic, which could be ruled out by extending the restriction on double topics to situations in which one of the topics is non-overt. It has been shown elsewhere, however, that zero objects in Tuvaluan share many properties with zero pronouns (Besnier 1983); thus, the Government-Binding account outlined above, however attractive, is problematic.

10. Unlike their Tuvaluan counterparts, as will be seen below, Niuean TS-ergative verbs do not appear to form a semantically-defined class.

11. The following list exhausts all maua-type verbs attested to date: maua *can, able to*, iloa *(mentally) able to, skilled at*, fai *to mean to, to act in order that*, STV masaua and NTV manatua *to remember*, STV kkafi *(physically) able to, strong enough to*, and mafai *capable of/to*.

12. Closer to home, Rapanui (Easter Island) is analysed by Alexander (1981) as exhibiting a split intransitive, or active, case-marking pattern, in which the morpheme cognate to the ergative case preposition in Western Polynesian languages denotes volitional subjecthood, while the non-overt case marking of a subject, whether transitive or intransitive, denotes non-volitionality. Coincidentally, Rapanui also exhibits a certain amount of word-order freedom (Alexander 1981), although it appears to be underlyingly verb initial (Chapin 1978).

13. A sample count over a 508-clause corpus of informal, non-elicited conversational discourse yielded, from a total of 300 main or independent clauses, only 14 (5%) main or independent clauses with two non-oblique nominal arguments.

14. In this type of structure, the possessive marker on the object does not denote any possessor-possessed relationship between the semantic subject and the object. In sentence (23), for instance, the fish may have been someone else's as soon as it was caught. Furthermore, the encoding of an alienable-inalienable distinction in Tuvaluan possessive morphology (see Besnier 1981b) allows a high vs low agentivity distinction to be marked in these structures.

15. The same conclusion is suggested informally by Chung (1978):

> It is tempting to suggest that [the] discrepancy [between the basic word order and the most frequent word order in Kapingamarangi] reflects an incipient word order change, and Kapingamarangi may eventually reanalyse its most frequent word order - SVO - as the basic word order. Similar changes may well have occurred in other Outlier languages, which according to Clark (1976) have SVO as their only surface word order. (1978:20)

16. Explanations for language change that invoke relational ambiguity have also been criticised by Li and Thompson (1974:211) and Moravcsik (1978), amongst others.

17. It is also likely that the West Uvean system illustrated in (33) is also the result of influence from Iaai, and not an innovation shared with Tuvaluan, Tikopia, Anutan, etc.

APPENDIX

Abbreviations

ANU	Anutan
FUN	Funaafuti
LUA	Luangiua
MEF	Mele-Fila
NEA	Nanumea
NGA	Nanumaga
NKF	Nukufetau
NKL	Nukulaelae
NTO	Niutao
NTV	Northern Tuvaluan
STV	Southern Tuvaluan
TIK	Tikopia
TUV	Tuvaluan
WFU	West Futunan
WUV	West Uvean

Art	article
Cmp	complementiser
Cnt	contrastive marker
Dxs	deictic adverb
Erg	ergative case
Foc	focus marker
Inc	inchoative
Nps	non-past
Num	numeral marker
Prf	perfective
Pst	past
Sbj	subjunctive conjunction
Spc	specific
Trn	transitivising suffix
2	dual
3	plural
+	morpheme boundary

REFERENCES

ALEXANDER, James D.

1981 Case-marking and passivity in Easter Island Polynesian. *Oceanic Linguistics* 20:131-149.

ANDERSON, J.M. and C. JONES, eds

1974 *Historical linguistics*, vol.1: *Syntax, morphology, internal and comparative reconstruction*. North Holland Linguistic Series 12a. Amsterdam: North-Holland.

BAYARD, Donn T.

1967 *The cultural relationships of the Polynesian Outliers*. Otago University Studies in Prehistoric Anthropology Series 9. Dunedin: Department of Anthropology, University of Otago.

BESNIER, Niko

1981a Transitivity and case in Tuvaluan. M.A. Paper, Department of Linguistics, Stanford University.

1981b *Ttou tauloto te ggana tuuvalu: a course in the Tuvaluan language*. Funafuti: U.S. Peace Corps.

1983 Empty pronouns in Tuvaluan. Typescript, Department of Linguistics, University of Southern California.

1985 The local organisation of zero anaphora in Tuvaluan conversation. *Te Reo* 28:119-147.

1986 A dialectal survey of Tuvalu. To appear in *PL*.

BIERWISCH, Manfred

1963 *Grammatik des deutschen Verbs*. Studia Grammatica 2. Berlin: Akademierverlag.

CHAPIN, Paul G.

1978 Easter Island: a characteristic VSO language. In Lehmann, ed. 1978: 139-168.

CHOMSKY, Noam

1981 *Lectures on government and binding*. Studies in Generative Grammar Series 9. Dordrecht: Foris.

CHUNG, Sandra L.

1978 *Case marking and grammatical relations in Polynesian*. Austin: University of Texas Press.

CLARK, Ross

1975a *Mele notes*. Working Papers in Anthropology 40. Auckland: Department of Anthropology, University of Auckland.

1975b *Two Mele texts*. Working Papers in Anthropology 44. Auckland: Department of Anthropology, University of Auckland.

1976 *Aspects of Proto-Polynesian syntax*. Te Reo Monographs. Auckland: Linguistic Society of New Zealand.

1977 Eastern Oceanic transitive suffixes and the genesis of rules. Paper presented at the Austronesian Symposium, LSA Summer Institute, University of Hawaii.

1978 The New Hebridean Outliers. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, fascicle 2, 911-928. *PL*, C-61.

1986 Linguistic convergence in Central Vanuatu. In *FOCAL II*, 333-342.

COLE, Peter

1974 Indefiniteness and anaphoricity. *Language* 50:665-674.

COMRIE, Bernard

1981 *Language universals and linguistic typology*. Chicago: University of Chicago Press.

COWELL, Reid

1951 *The structure of Gilbertese*. Beru: Rongorongo Press.

CUMMING, Susanna

1986 Word order change in Malay. In this volume, 97-111.

DOUGHERTY, Janet W.D.

1983 *West Futuna-Aniwa: an introduction to a Polynesian Outlier language*. University of California Publications in Linguistics, vol.102. Berkeley: University of California Press.

DURANTI, Alessandro

1981 *The Samoan* fono*: a sociolinguistic study*. *PL*, B-80.

DURANTI, Alessandro and Elinor OCHS

1979 Left-dislocation in Italian conversation. In Talmy Givón, ed. *Syntax and semantics*, vol.12: *Discourse and syntax*, 377-416. New York, Academic Press.

EARLY, Robert J.

1981 Minor morphemes in the language of Tikopia. M.A. thesis, Department of Anthropology, University of Auckland.

FEINBERG, Richard

n.d. Collection of Anutan oral traditions. Typescript.

FINNEY, Joseph C.

1983 Nanumanga verbal syntax: a study in variation, with implications for Proto-Polynesian. *Journal of the Polynesian Society* 92:107-115.

GEORGOPOULOS, Carol

1986 Palauan as a VOS language. In this volume, 187-198.

GERAGHTY, Paul A.

1978 Fijian dialect diversity and foreigner talk: the evidence of pre-missionary manuscripts. In Schütz, ed. 1978:51-67.

1983 *The history of the Fijian languages*. Oceanic Linguistics Special Publication 19. Honolulu: University Press of Hawaii.

GIVÓN, Talmy

1977 The drift from VSO to SVO in Biblical Hebrew. In Li, ed. 1977:181-254.

1979 *On understanding grammar*. Perspectives in Neurolinguistics and Psycholinguistics Series. New York: Academic Press.

GREENBERG, Joseph H.

1966 Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg, ed. *Universals of language*, 73-113. 2nd edn. Cambridge, Mass.: MIT Press.

HARLOW, Raymond

1986 The actor emphatic construction of the Eastern Polynesian languages. In this volume, 297-308.

HARRIS, Alice C.

1982 Georgian and the unaccusative hypothesis. *Language* 58:290-306.

HAWKINS, John A.

1983 *Word order universals*. Quantitative Analyses of Linguistic Structure Series. New York: Academic Press.

HOOPER, Robin

1986 Word-order variation in Samoic-Outlier languages. In this volume, 269-295.

HOPPER, Paul J. and Sandra A. THOMPSON

1980 Transitivity in grammar and discourse. *Language* 56:251-299.

HOWARD, Irwin J.

1981 Proto-Ellicean. In Jim Hollyman and Andrew Pawley, eds *Studies in Pacific languages & cultures in honour of Bruce Biggs*, 101-118. Te Reo Monographs. Auckland: Linguistic Society of New Zealand.

HUANG, G.-T. James

1983 On the distribution and reference of empty pronouns. Typescript. National University of Taiwan.

HYMAN, Larry M.

1975 On the change from SOV to SVO: evidence from Niger-Congo. In Li, ed. 1975:113-147.

KEENAN, Edward L.

1976 Towards a universal definition of 'subject'. In Li, ed. 1976:305-333.

1978 The syntax of subject-final languages. In Lehmann, ed. 1978:267-327.

1979 On surface form and logical form. *Studies in the Linguistic Sciences* 8/2:1-41.

KOSTER, Jan

1975 Dutch as an SOV language. *Linguistic Analysis* 1:111-136.

LAKOFF, George

1968 *Pronouns and reference.* Bloomington: Indiana University Linguistic Club.

LEHMANN, Winifred P.

1973 A structural principle of language and its applications. *Language* 49:47-66.

LEHMANN, Winifred P., ed.

1978 *Syntactic typology: studies in the phenomenology of language.* Austin: University of Texas Press.

LEVIN, Juliette and Diane MASSAM

1986 Classification of Niuean verbs: notes on case. In this volume, 231-244.

LI, Charles N., ed.

1975 *Word order and word order change.* Austin: University of Texas Press.

1976 *Subject and topic.* New York: Academic Press.

1977 *Mechanisms of syntactic change.* Austin: University of Texas Press.

LI, Charles N. and Sandra A. THOMPSON

1974 Historical change of word order: a case study in Chinese and its implications. In Anderson and Jones, eds 1974:199-217.

1976 Subject and topic: a new typology of language. In Li, ed. 1976: 457-489.

MALLINSON, Graham and Barry J. BLAKE

1981 *Language typology: cross-linguistic studies in syntax.* North-Holland Linguistic Series 46. Amsterdam: North-Holland.

MERLAN, Francesca

1985 Split intransitivity: functional oppositions in intransitive inflection. In Johanna Nichols and Anthony C. Woodbury, eds *Grammar inside and outside the clause: field linguists look at current issues,* 324-362. Cambridge: Cambridge University Press.

MOAG, Rodney F.

1978 Standardization in Pidgin Fijian: implications for the theory of pidginization. In Schütz, ed. 1978:69-81.

MORAVCSIK, Edith

1978 On the limits of subject-object ambiguity tolerance. *Papers in Linguistics* 11:255-259.

MOYSE-FAURIE, Claire and Françoise OZANNE-RIVIERRE

1983 Subject case markers and word order in New Caledonia and Loyalty Islands languages. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.4: *Thematic variation*, 113-152. *PL*, C-77.

OCHS, Elinor

1982 Ergativity and word order in Samoan child language. *Language* 58: 646-671.

PARKER, F.

1980 Typology and word order change. *Linguistics* 18:269-288.

PAWLEY, Andrew

1967 The relationships of the Polynesian Outlier languages. *Journal of the Polynesian Society* 76:259-296.

REEDY, Tamati

1977 The *-Cia* suffix in Takuu. Typescript, Department of Linguistics, University of Hawaii.

SALMOND, Anne

1974 *A generative syntax of Luangiua, a Polynesian language*. Janua Linguarum Series Practica 152. The Hague: Mouton.

SCHÜTZ, Albert J., ed.

1978 *Fijian language studies: borrowings and pidginization*. Bulletin of the Fiji Museum 4. Suva: Fiji Museum.

SMITH, Norval V.

1981 Consistency, markedness and language change: on the notion 'consistent language'. *Journal of Linguistics* 17:39-54.

TRUSSEL, Steve

1979 *Kiribati (Gilbertese) grammar handbook*. Peace Corps Language Handbook Series. Brattleboro, Vermont: The Experiment in International Living.

VENNEMANN, Theo

1974 Topic, subject, and word order: from SXV to SVX via TVS. In Anderson and Jones, eds, 1974:339-376.

1975 An explanation of drift. In Li, ed. 1975:269-305.

WANG, Judith

1976 The article *a* in Vaitupu. *Working Papers in Linguistics, University of Hawaii* 8/4:109-120.

DISCOURSE CONSTRAINTS ON WORD-ORDER VARIATION IN SAMOIC-OUTLIER LANGUAGES

Robin Hooper

Some Samoic-Outlier languages show frequent use of verb-medial word order, in addition to the VS(O) pattern reconstructed for PPN and found in most languages of the Polynesian family. At one extreme are Nukuoro and Kapingamarangi. Published texts in these languages show that when the noun phrase (NP) constituents are not suppressed, one of them, usually the subject, precedes the verb. Some investigators have noted a similar tendency in other languages (for example, Early 1981 for Tikopia, Besnier 1981 for Tuvalu). This situation has led to speculations on the possibility of a change in progress from verb-initial to subject-initial word order (Chung 1978:20, Early 1981:118).

Clark (1976:39-40) proposes that the historical source of subject-initial order is the topicalised sentence common throughout Polynesia, and that loss of the topic marker ko before the fronted NP in some languages has introduced a formal contrast between the once identical topic and focus constructions. The correctness of this proposal will be assumed in what follows, and is supported by the situation in some languages in which the use of ko on topics is somewhat unstable.

The process of ko-fronting, in those languages which retain it, relates pairs of sentences like the following (all examples in this initial section are in Tokelauan):

(1)a. E talanoa ki māua i te fakakilipati
T/A *talk* *we-2* *in the Gilbertese*

b. Ko ki māua e talanoa i te fakakilipati
top. *we-2* T/A *talk* *in the Gilbertese*

He and I converse in Gilbertese. (Tioita 3)

Such ko-fronted sentences are exceedingly common, and although they are usually considered to be, syntactically, a marked variant of the basic verb-initial sentence type, I do not regard them as in any way marked or emphatic in a pragmatic sense. They may be compared with another type of ko-fronted topic construction which could be more aptly called left-dislocation. In this sentence type, there is a noticeable intonation break after the fronted NP, which may be quite tenuously related syntactically to the rest of the sentence:

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 269-295. *Pacific Linguistics*, C-93, 1986.

(2) Ko te ulugā fonu, kua i a te koe lava te tonu
top. *the pair turtle* T/A *at* art. *you* int. *the decision*
As for a pair of mating turtles, the decision rests entirely with you.
(*Tokelau fishing* 4)

These two types of ko-fronted topic can co-occur in the same clause, as in (10. 21-22) below.

Ko-fronting can also be used to indicate a focused NP. Clark (1976:37) comments on the ambiguity of examples like (1b) between topic and focus, so that (1b) could in theory at least be glossed *It is we two who converse in Gilbertese.* In practice however, NPs which are focused in this way are distinguished by additional forms of emphasis. They are usually followed by an intensive particle such as Tokelauan lava, and in addition a resumptive pronoun usually appears in the clause, even in cases of focused subjects, which do not always require anaphora when topicalised:

(3) Ko koe lava te tautai e kē faia
foc. *you* int. *the captain* T/A *you do*
It is you, the master fisherman, who does this. (*Tokelau fishing* 3)

Closely resembling this construction is another type of ko-fronted focus construction which contains a relative clause with deleted head NP. This structure, common in Nuclear Polynesian languages (Clark 1976:62-63), is the type normally proferred by Tokelauan informants in response to translation requests for English cleft sentences. It is also the form taken by some 'who?' and 'what?' questions:

(4) Ko ai te kua fano?
pred. *who the* T/A *go*
Who has gone?

(5) He ā te tagi ai koe?
a what the weep pro. *you*
What are you crying about?

(6) Ko nā tamaiti te na olo
pred. *the children the* T/A *go*-pl.
It was the children who went.

While left-dislocated and focused NPs constitute a proportion of the sentence-initial NPs in the texts examined below, by far the majority are simple topics of the kind exemplified in (1b).

Sentence-initial position is available to any NP, although requirements for post-verbal anaphora differ according to the syntactic function of the fronted NP. In certain types of text, semantic patients, obliques, and genitives are fronted almost as often as subjects. For this reason I will not talk about VS(O) as opposed to SV(O) word order, as I consider this an incorrect formulation of the problem. I will talk instead of verb-initial and NP-initial clauses. My concern in this paper is to investigate the extent to which NP-initial clauses occur in a number of languages, and to specify the types of context in which they occur.[1]

This demands some examination of the syntax of narrative. Recent literature on narrative structure and verb aspect distinguishes two types of clause in narnative, 'narrative' or 'foreground' clauses, and 'background' clauses. This

distinction was developed in Forsyth 1970 and Labov 1972, and has been extended to a number of languages in recent work on foregrounding and backgrounding in oral narrative (e.g. Hopper 1979a, 1979b, 1982, Schiffrin 1981, Silva-Corvalán 1983).

In any story, whether a traditional tale, a narrative of personal experience or an historical account, there are certain clauses which relate the successive events of the story. They occur, iconically, in the same sequence as the events to which they refer; if two such clauses are transposed, the order of events is also understood to be transposed. These are the narrative clauses. Stories also contain a great deal of supportive and explanatory material which does not in itself advance the action, and this material is expressed in background clauses - relative clauses, temporal clauses, any clauses which refer to situations which obtained prior to the time of the narrative, or which extend over a part or the whole of the time the narrative is in progress. These clauses can be moved around to a greater or lesser extent without doing violence to the story line. Background clauses can be classified into a number of types according to their function in the narrative as a whole. Here I will simply point out that a number of background clauses usually occur at the beginning of a story, giving preliminary information about characters and setting, and possibly summarising the point of what is to follow. Others occur during the course of the story, conveying information which the listener must have in order to understand the course of events correctly. Some background clauses are evaluative, in that they convey obliquely or directly the narrator's feelings about the events he relates.

Briefly, we can say that narrative clauses have the following characteristics: they are positive and declarative, and refer to dynamic situations in strict chronological sequence; it is common for sequences of narrative clauses to have the same subject, which is thus old information and is presupposed, whereas the verb and its arguments constitute the new information. Background clauses, on the other hand, frequently refer to stative and durative situations, which may be prior to, concurrent with or future to the time of the narrative, and they may be negative or irrealis.

It has been shown for a number of languages that narrative clauses differ syntactically from background clauses. French and Russian use tense-aspect marking, Old English used word order, and Malay uses word order and verb morphology, to mark the distinction (Hopper 1979a, 1982). Polynesian languages, as we shall see, use both tense-aspect morphology and word order.

On the subject of languages which rely on word-order strategies, Hopper writes as follows:

> In this strategy, it is the position of the verb which is crucial. The verb is the location of new, narrative-advancing information. The verb's complement may or may not contribute to the narrative, and the subject of the verb is least likely to play a significant role in the story-line. Thus we typically find the verb in one of the two most prominent positions in the clause, the beginning or the end, and the subject of the verb is likely to be highly presuppositional, in fact, usually identical with the preceding subject. (Hopper 1979a:240)

Elsewhere he says

> Word order changes when a dislocation becomes grammaticalized and loses its marked value, e.g. when the subject NP in a VSO language becomes topicalized and placed before the verb. In other words, the higher degree of flexibility in an imperfective (non-aorist) sentence type is more conducive to the development and fixing of new word order types.
>
> (Hopper 1979b:51)

In order to consider the extent to which these generalisations apply to Samoic-Outlier languages, I consider first tense-aspect and word-order patterns in narrative and expository texts in three languages, Tokelauan, East Futunan, and Rennell. Unless otherwise indicated, all the texts cited are transcriptions of spontaneous oral discourse.

In the oral narratives of these three languages, narrative and background clauses are distinguished in the two ways referred to above. Of these, the most consistent, though not the most significant for our present purposes, is tense-aspect marking. Almost all narrative clauses are unmarked for tense-aspect. The bare verb stem indicates perfectivity and sequencing. A sequence of clauses with unmarked verbs indicates a sequence of events in perfective aspect, i.e. viewed as single events without regard to the duration of the act or its internal phases. There are two types of exception to the use of the unmarked verb in narrative clauses. In narratives of personal experience, it is common for a past tense marker (e.g. Tokelauan and East Futunan na) to be used sporadically, especially in the initial section of the narrative. Once past time reference is established, unmarked verbs take over. The second exception is the use of a reflex of PPN *kua (usually referred to as a perfect marker in accordance with one of its main functions) as an intensifier in some clauses. This use of kua resembles the historic present alternation common in English and other languages (see Wolfson 1979, Silva-Corvalán 1983), and must be distinguished from its use in background clauses to mark anteriority or state. The common conjunctions Tokelauan oi, East Futunan and Rennell o *and* (sequential), East Futunan ti *and, then*, and Tokelauan and Rennell kae *and, but*, are all used in narrative clauses.

Background clauses are of two types. They are often equational in structure, in which case they have nominal predicates. If they are verbal, they almost invariably contain a tense-aspect marker, which indicates the temporal relation between the situation to which they refer and the time of the narrative event.

The second parameter is that of word order. Narrative clauses are verb initial; the bare verb stem can in fact only occur sentence initially, or in second position after a temporal adverb, as in example (8.1) below. In background clauses, word order is far more flexible, and NP-initial order is common. Nominal predicate clauses are by definition NP initial, but a great many verbal background clauses are also NP initial as a result of the topicalisation processes described above. In folk tales, background material is comparatively sparse, and confined to what is necessary for understanding the narrative, whereas in narratives of personal experience, evaluative and other background material, including direct speech, can amount to up to half of all clauses, providing a rich mine of NP-initial clauses. These frequently occur after perception verbs and verbs of saying.

In the extracts given below, I follow the convention of starting the narrative clauses at the left-hand margin, and indenting all background clauses. Note that (7.1-7) is the introductory or orientation section of a tale. Six of

these clauses are NP initial. Note also at (7.13) the use of a repeated unmarked verb. This device is quite commonly used to indicate a repetitive, durative, or habitual situation which is seen as a single episode for the purposes of the narrative.

(7) 1 Ko te Kimoa e nofo i te Kimokimoā,
top. *the Rat* T/A *live at the K.*
Rat lived at Kimokimoā,

2 ko Āfinemata e nofo i Vaō.
top. *A.* T/A *live at V.*
Āfinemata lived at Vaō.

3 Ko te vāitaimi tēnā e pakū foki ki te fakaogeoge o nā fenua.
top. *the time that* T/A *fall too to the famine of the lands*
That period of time turned into a famine of the lands.

4 Ko te fenua o Āfinemata, e hēai he tino e fano ki ei.
top. *the land of A.* T/A *not a person* T/A *go to there*
As for Āfinemata's land, nobody went there.

5 Ko Āfinemata he fafine hāuā.
top. *A.* *a woman savage*
Āfinemata was a ferocious woman.

6 E maua he tino e Āfinemata, oi kave tuku ki he lalo fatu,
T/A *catch a person by A.* cnj. *take put to a under stone*

tuku ki he tafito lākau.
put to a base tree
When Āfinemata caught a person, she took him and put him under a stone or put him at the base of a tree.

7 E vēnā te uiga ....
T/A *like-that the nature*
That's what she was like ....

8 Ko te fale o Āfinemata hōvē e i te Lotolahi.
top. *the house of A.* *perhaps* T/A *at the L.*
Āfinemata's house was probably at Lotolahi.

9 Hau ai te Kimoa, ki tua ki gatai, ki tua ki gatai.
come pro. *the Rat* *to ocean-side to lagoon*
Rat came along, from ocean side to lagoon side, from ocean to lagoon.

10 Kikila mai te Kimoa
look dir. *the Rat*
Rat looked out,

11 ko Āfinemata e tae otaota
top. *A.* T/A *clear up rubbish*
Āfinemata was picking up rubbish

12 Hau te Kimoa
come the Rat
Rat came,

13 taukalo, taukalo, taukalo
dodge
he dodged and dodged and dodged,

14 kake ki luga o te niu
climb to above of the coconut-palm
he climbed up the coconut palm,

15 fano ki luga o te niu
go to above of the coconut-palm
went to the top of the coconut palm.

16 Fakatatau ifo lava e te Kimoa te mea
judge down int. *by the Rat the thing*
Rat judged carefully down to the place

17 e i ei ia Āfinemata,
T/A *at there* art. *A.*
where Āfinemata was,

18 kati te puakoile
bite the forming-coconut
bit the forming coconut,

19 togi.
throw
and threw it.

20 E hē tokaga lele ia Āfinemata.
T/A *not pay-attention* int. art. *A.*
Āfinemata paid no attention.

(Huntsman 1977:26,28)

The next two examples are from narratives of personal experience.

(8) 1 Te lua o aho, kamata loa toku havalivali mai ki te kakai.
the second of day begin then my walk-R. dir. *to the village*
On the second day, I took my first walk to the village.

2 Fakatau taku moli tākele i te aho muamua,
buy my soap bathe on the day first
I bought a cake of soap on the first day /i.e. of the walks/

3 fakatau taku apa ika i te lua o aho.
buy my tin fish on the second of day
and I bought a can of fish on the second day.

4 Hēai he mea e kiki ai.
not a thing T/A *eat-with* pro.
I had no starch food to accompany it.

5 Na maua lā taku popo i toku havalivaliga.
T/A *obtain* int. *my coconut on my walking*
However I had obtained a coconut while I was walking.

6 Ia. Ko te popo lā tēnā ma te apa ika na ola ai au.
yes top. *the coconut* int. *that and the tin fish* T/A *live* pro. *I*
Yes, that coconut and the tin of fish were what I survived on.

7 Fano au i tētahi vāitaimi
go I on another time
I went on another occasion

8 fakatau mai taku hikaleti.
buy dir. *my cigarettes*
and bought some cigarettes.

9 Ko te taimi lā ia na iloa ai au.
foc. *the time* int. *that* T/A *know* pro. *I*
It was on that occasion that I was found out.

10 Fehili mai ai te tamāloa -
ask dir. pro. *the fellow*
The young man asked me -

11 ko Eneleo te igoa o te fakatau koloa -
pred. *Henry the name of the sell goods*
the shop-keeper's name was Henry -

12 fehili mai ki a te au
ask dir. *to* art. *I*
he asked me

13 pe ko au na hau i fea.
Q. top. *I* T/A *come from where*
where I had come from.

14 Manatua lā taku fakakupu -
remember int. *my speech*
I remembered my speech -

15 oi oti lā te igoa kua kō māua i kinā, ko Lalomānu.
conj. *only* int. *the name* T/A *I hear at there,* pred. *L.*
there was only one name I knew of round there, which was Lalomānu -

16 fano loa taku fakakupu
go then my speech
my word went out:

17 "Ko Lalomānu, ko au na hau i Lalomānu."
pred. *L.* top. *I* T/A *come from L.*
"Lalomānu, I came from Lalomānu."

18 "I te kāiga o ai" ko te fehili ia a Eneleo.
from the family of who pred. *the question that of H.*
"From whose family?" that was Henry's question.

19 Uma loa toku malamalama ma hoku fia tali fakapepelo atu.
finished then my knowledge and my wish reply deceive dir.
Both my knowledge and my wish to deceive came to a halt.

20 la au e mataku nā ia au e toe kavea ki luga o te vaka.
art. *I* T/A *fear lest* art. *I* T/A *again taken to above of the ship*
I was scared that I might be taken on board the ship again.

21 Fakahako loa e au
set-straight then by me
I told him straight,

22 ko au na hau mai i te vaka ika e i Apia
top. *I* T/A *come* dir. *from the boat fish* T/A *at A.*
that I had come from the fishing vessel which was at Apia,

23 ko au na hola
top. *I* T/A *run away*
that I had deserted

24 ko au kua muhu fano i te vaka.
top. *I* T/A *unwilling go on the boat*
and that I was unwilling to go on the ship.

(*Hola* 10-11)

(9) 1 Fakatū loa lava te lā,
set up then int. *the sail*
Then we set up the sail

2 ka ko au e i te ama.
conj. top. *I* T/A *on the outrigger*
while I was on the outrigger.

3 Tū loa lava te lā,
stand then int. *the sail*
The sail went up,

4 tele ifo
go-swiftly dir.
we sailed downwind

5 āgai ki Fenuafala.
head for to F.
and made for Fenuafala.

6 Ko toku pakūga ifo lava tēnā ki luga o te tataoga o te
top. *my falling* dir. int. *that to above of the weight of the*

ama, moe, e hē kō iloa he mea.
outrigger sleep T/A *not I know a thing*
As for that collapsing down of mine on top of the outrigger platform,
I fell asleep and was not aware of a thing.

7 Pe te galu e fafati mai
whether the wave T/A *break* dir.
If a wave broke over me,

8 e hē kō iloa he mea.
T/A *not I know a thing*
I wasn't aware of a thing.

9 Kae malama ake te taeao,
conj. *dawn* dir. *the morning*
Well, the next day dawned,

10 pupula ake au,
wake dir. *I*
I woke up,

11 ko te vaka e tautau i te ava.
top. *the canoe* T/A *anchor in the pass*
the canoe was anchored in the pass.

12 Ko Teao ma Patemo nae hī malau i te pō.
top. *T. and P.* T/A *line-fish soldier-fish in the night*
Teao and Patemo had been fishing for soldier-fish during the night.

13 Hōvē ko ki mātou kua leva te pā atu ki tua
perhaps top. *we* T/A *long ago the arrive* dir. *to ocean-side*
o Fenuafala
of F.
It seemed it was a long time since we had arrived offshore of Fenuafala.

14 Ko nā aho iēnā kō, Hapeta e failele.
top. *the days those* int. *H.* T/A *pregnant*
At that time, you know, Hapeta was pregnant.

15 Olo ai ki mātou
go pro. *we*
We set off,

16 tiaki te tui puhi i kinā.
throw away the string of eels at there
and we threw away the string of eels at that point.

(*Tala faiva* 9-11)

The East Futunan example below contains tense-aspect marking in the direct speech passage, and in the relative clause at line 6 and the stative clause at line 16. Note the double topic constructions in lines 21 and 22. Narrative clauses are unmarked for tense-aspect but contain the conjunctions ti *and, then* and o 'sequential'.

(10) 1 Aga ake tama
face dir. *young man*
The young man set to,

2 tae a taulua vai masa
take two water bottle empty
and got two empty water bottles

3 o ano mo ia
conj. *go with him*
and took them with him

4 o tausa'ele i le ala ki 'uta la.
conj. *walk along on the road to inland then*
and set out along the road to the interior.

5 Ti kaku aia ki Le Puna
then reach he to L.P.
Then he arrived at Le Puna

6 na pulou ai le matu'a
T/A *cover head* pro. *the parent*
where the parent had been crowned /with bark-cloth/,

7 ti aga aia
conj. *face he*
and he went

8 o to'o le siapo
conj. *take the bark cloth*
and took the piece of bark cloth

9 o 'efi,
conj. *carry under arm*
and put it under his arm

10 ti puna mai aia ki tai.
conj. *flee* dir. *he to coast*
and he fled towards the coast.

11 Au le temonio mei muli
come the demon from behind
The demon came after him ....

12 o kaku ki tai
conj. *reach to coast*
and they arrived at the coast,

13 o siga mate le tama ki fale
conj. *fall dead the young man to house*
and the young man fell as if dead in the house.

14 Ti au le temonio
conj. *come the demon*
Then the demon came,

15 o tu'u mei fafo
conj. *stand outside*
and stood outside.

16 Kua i le fenua
T/A *afraid the people*
The people were afraid.

17 o pati mai ki lona makopuna,
conj. *say* dir. *to his grandson*
And he said to his grandson,

18 "Ko koe, le'ai se kē matu'a
top. *you not not you grow up*
"As for you, you will not grow

19 "ti e le'ai se kē malu
conj. T/A *not not you sheltered*
"and you will not prosper.

20 "E kē mele,
T/A *you offensive*
"You are offensive,

21 "ti ko le tu'a o lou fenua, ko vai e ifo ki tu'a,
conj. top. *the back of your land* top. *water* T/A *go down to back*
"and as for the back of your land, water will fall to the back,

22 "ko se la'a, ko koe fa'i e 'ulu'aki masa."
top. *a sun* top. *you* int. T/A *be first dry*
"and in a drought, you will be the first to run dry."

23 Ti puli le temonio.
conj. *disappear the demon*
Then the demon disappeared.

(*Talatuko o sigave* 3-4)

The Rennell and Bellona traditional tales collected in Elbert and Monberg 1965 tend to contain little background material except for the opening clause(s), which are regularly NP initial. Two long narratives of experience, at the end of the collection, contain a great deal of background and evaluative material, including direct speech. An extract from one of them is given below:

(11) 1 Hano au ki te tokagua hakahua
go I to the two chiefs
I went to the two chiefs

2 o hakakegeu kinai
conj. *speak to*-pro.
and spoke to them

3 hai atu kinai au,
say dir. *to*-pro. *I*
I said to them,

4 "Gu atatou 'atua e sia hai kau taa'ia."
two our gods T/A *want do I destroy*
"I want to destroy our two gods."

5 Ka gua hakahua noko matataku, noko hai mai, "Mano he'e 'aonga."
conj. *two chiefs* T/A *afraid* T/A *say* dir. *maybe not necessary*
But the two chiefs were afraid and said "Maybe better not."

6 O hai atu au,
conj. *say* dir. *I*
And I said,

7 "Ko au e tu'uganga i te to'a o God."
top. *I* T/A *rely on the power of G.*
"I rely on the power of God."

8 loo mugi o mangangao.
then after conj. *willing*
Then they agreed.

9 O kakabe e kigaaua ia te au,
conj. *take by they*-2 art. *I*
And they took me,

10 o boo
conj. *go*-pl.
and went

11 o tuiaki kia te au,
conj. *guide to* art. *I*
and guided me.

12 Namaa hetae kinai kimatou,
when arrive to-pro. *we*
When we arrived there

13 hakaputu kimatou
gather together we
we gathered close together,

14 o sogi ki te 'Aitu
conj. *pray to the Lord*
and prayed to the Lord

15 e i te gangi,
T/A *in the sky*
who is in heaven,

16 kae he'agiko gua hakahua i te mouku,
conj. *run away two chiefs in the bush*
but then the two chiefs ran away to the bush

17 i te me'a gaa noko matataku ki gua 'atua
because int. T/A *fear to two gods*
because they were afraid of the two gods

18 kaa taa'ia.
T/A *destroy*
who were to be destroyed.

19 Namaa 'oti taku sogi ki te 'Aitu i te gangi
when finished my pray to the Lord in the sky
When I finished my prayer to the Lord in heaven,

20 kae taa tohitohi e au i te 'aakisi
conj. *strike by me with the axe*
I broke /the gods/ to bits with my axe,

21 kae giu iho kimatou ki te hakatahinga,
conj. *turn* dir. *we to the gathering*
then we went back to the gathering

22 o ngengege atu kinai,
conj. *call* dir. *to*-pro.
and called out to it

23 "Gua 'atua kua taa'ia!"
two gods T/A *destroyed*
"The two gods have been destroyed!"

24 Na hu'ai'agiki noko he'ika'ika'aki toto'a.
the high chiefs T/A *angry very*
The important chiefs were very angry.

25 Ko au noko mataku ma'u kinai.
top. *I* T/A *afraid* int. *to*-pro.
I was thoroughly afraid of them.

(Elbert and Monberg 1965:235A,90-94)

I now want to examine some expository language. This kind of discourse typically contains a high proportion of NP-initial clauses. A 900-word East Futuna written text which describes in detail the geography of the area contains 50% NP-initial clauses. An extract is given below:

(12) Ti ko le 'ana o Lita e tu'u 'i tai i le matātai fela'aki'i mo
and top. *the cave of L.* T/A *stand by sea on the beach just like*

Vai-kinafa. Ko le 'ulu'aga o le 'ana o Lita e fai pē se matapā
V. top. *the entrance of the cave of L.* T/A *make like a door*

lasi o se ekelesia. Ti ko Vai-kinafa ko se 'ana lasi foki ti e
big of a church. and top. *V.* pred. *a cave big too and* T/A

fulumālie ai, o fai pē letasi o Sila. Ko le 'ana ko le
beautiful pro. conj. *make like the one of S.* top. *the cave* app. *the*

Tanoa e tu'u 'i le matātai 'ofi ki Loka, 'ana e 'iki'iki kā ko le
T. T/A *stand on the beach near to L. cave* T/A *small but* top. *the*

ne'a fai kese e ma'ua 'i loto o le gā ne'a e sali 'i ai le
thing different T/A *found at inside of the place* T/A *flow from* pro. *the*

vai. E fai pē se kumete, kā ko le fatu fulumālie ai, ti ko
water T/A *make like a bowl but* top. *the stone beautiful* pro. *and* top.

lona fatu 'i loto e gigila'ia pē se tilo'ata ma'uiga na fa'u
its stone at inside T/A *shine like a mirror meaning* T/A *constructed*

mo fakamālama o le 'ana.
for illumination of the cave.

The cave of Lita stands on the beach at the seaside just like Vai-kinafa. The entrance of the cave of Lita is constructed like a great church door. Vai-kinafa is also a big cave, and a beautiful one, and shaped like the one at Sila. The cave /called/ the Tanoa stands on the beach near Loka, a small cave, but an unusual feature is to be found inside the place from which the water flows. It is shaped like a bowl but beautiful stone, and the stone inside shines like a mirror, as though it were constructed for the illumination of the cave.

(*Fetuuaho* No.8:8)

In a large Tokelauan corpus of expository texts, composed of transcriptions of spontaneous and semi-spontaneous speech on the subject of fishing techniques and customs, and translations of medical information circulated by the local doctors, we find that NP-initial clauses range from one third to over one half of all clauses. There is no significant difference between the indigenous, spoken language of the fishing texts, and the foreign, translated, written language of the medical texts. Some of these NP-initial clauses are topicalisations, some are focused constructions, some are equational nominal sentences. Others are more complex structures based on the equational sentence and analogous to English cleft and pseudo-cleft sentences:

(13) Ko te mea muamua e tatau ona kē faia i te afiafi tēnā,
top. *the thing first* T/A *necessary* conj. *you do on the evening that*

ko koe ke kai ke mākona lelei lava.
top. *you* conj. *eat* conj. *satisfied well* int.

The most important thing that you have to do that evening, is that you must eat till you are thoroughly satisfied.

(*Tokelau fishing* 8)

Note that in most such cases the same idea could have been expressed in a verbal sentence. Language such as this, like the language of conversation, is concerned with people, things, opinions, facts, theories. Only occasionally will sequences of events be recounted; when they do occur they are encoded in verb-initial syntax, for example, lines 5-7 of the following extract.

(14) Ko tētahi tino e poapoa mai lava te pāla ki te tafāvaka,
top. *another person* T/A *cast bait* dir. int. *the wahoo to the side-canoe*

kae ko te kupu tū lava a te tautai ki te tino i te liu:
conj. top. *the word stand* int. *of the captain to the man at the bilge*

"Te ika nā matavakā!" Ko tona uiga o te kupu tēia, "Ko te
the fish cav. top. *its meaning of the word this* top. *the*

ika nahe poapoa maia ki te tafāvaka." Ko te foeliu
fish neg.imper. *cast bait* dir.-Cia *to the side-canoe* top. *the bilge paddler*

e kamata atu ai tona akoakoga. Kāfai kua fakatatau e te tautai
T/A *begin* dir. pro. *his learning when* T/A *judge by the captain*

kua lelei te tino tēnā i te foeliu, hiki atu loa ki te
T/A *good the man that at the bilge paddle shift* dir. *then to the*

hekelima ke nofo ai. Ko tona uiga, kā akoako ke mafai
fifth paddle conj. *sit* pro. top. *its meaning* T/A *learn* conj. *able*

ona kavea ia ma foemua. I te taimi kua teka atu ai te
conj. *become he as fore paddler at the time* T/A *leave* dir. pro. *the*

foeliu ki te hekelima, ko te tautai foki kua toe tōfia
bilge paddler to the fifth paddle top. *the captain too* T/A *again choose*

he tino i te tokalua tino iēnā i mua, ke hau ki loto i te
a man at the two men those at front conj. *come to inside at the*

foeliu, ke kamata atu ai ana āoga, auā ko te tino kua
bilge paddle conj. *begin* dir. pro. *his learning because* top. *the man* T/A

teka atu mai kinā, kua poto lava ia i nā mea uma.
leave dir. *from there* T/A *skilled* int. *he in the things all*

Another man /on first becoming bilge-paddler/ will lure the wahoo right up to the side of the canoe. Then the traditional saying of the captain to the man at bilge paddle is "Let not the fish matavakā*." The meaning of this expression is "Don't lure the fish up to the side of the canoe". So the bilge-paddler begins his apprenticeship. When the captain judges that this man is competent as bilge-paddler, he shifts him to fifth paddle to sit there. The significance of this is that he is going to study so that he can become a fore-paddler. At the time that the bilge-paddler moves to fifth paddle, the captain once again chooses a man from those two up front, to come in to bilge paddle to begin his learning, because the man who has just left that position is skilled at everything he has to do.*

(*Tokelau tautai* 5-6)

(15) Kae ko au e fakaali atu lava e au te mea e mautinoa e au.
conj. top. *I* T/A *show* dir. int. *by me the thing* T/A *certain by me*

Ko te mea kua fai foki e au. E hēai ni māhina i a te au, kae
top. *the thing* T/A *do also by me* T/A *not some moons to* art. *me* conj.

ko he itūmatagi. Ke koutou manatuagia te matagi tēia ... pe
top. *a wind direction let you*-pl. *remember the wind that whether*

ko te Taumuliava, ka ko koe e i luga i te utua. Ke kē
top. *the T.* conj. top. *you* T/A *at above at the point let you*

manatuagia, ko te vaka e tau ki te mata o te matagi. Kae
remember top. *the canoe* T/A *anchor to the eye of the wind* conj.

ke kē manatua foki, ko te mulivaka e fufuli tonu lava ki te
let you remember too top. *the canoestern* T/A *turn straight* int. *to the*

mea e oho mai ai te lā. E pōuli lā te mulivaka. E pōuli
thing T/A *rise* dir. pro. *the sun* T/A *dark* int. *the stern* T/A *dark*

te malae o te tautai i te tī lā. Tēnā lā e lea atu ai au.
the seat of the captain in the glare sun that int. T/A *say* dir. pro. *I*

I a te au, i te mea kua malamalama ai au ma te mea e kō
at art. *me in the thing* T/A *enlightened* pro. *I and the thing* T/A *I*

mautinoa. Ko te matāmatagi tēnā, kua tau te muāvaka ki te
certain top. *the wind direction that* T/A *align the bow to the*

matagi, ka ko te mulivaka e liliu tonu lava ki te tī lā.
wind conj. top. *the stern* T/A *turn straight* int. *to the glare sun*

Kua pōuli lā te malae o te tautai. Ko au la e hē tataoa
T/A *dark* int. *the space of the captain* top. *I* int. T/A *not lay-the-net*

ki te itū ama. Kae ko te mea aku e fai, e hē toe
to the side outrigger conj. top. *the thing my* T/A *do* T/A *not again*

taua te vaka ki te matagi, ka kua tau fakapala. Ko tēnā
align the canoe to the wind conj. T/A *align across wind* pred. *that*

te faiga. Kae ko iētahi hukehukega e i a te koe ma ni ō
the doing conj. top. *other questions* T/A *at* art. *you and some your*

māfaufauga. Ko te tamā mea tēnā e fia tali pukupuku atu ai
thinking top. *the small thing that* T/A *wish reply short* dir. pro.

au ki te tautai tēia.
I to the captain that

But as for me, I shall explain something of which I am certain. Something also which I have done. Months of the year are of no importance to me, but rather the direction of wind. Remember that wind ... perhaps it's the Taumuliava. Well, you are on the point. Now remember that the canoe is turned into the eye of the wind. And remember also that the stern is turned straight towards the direction in which the sun rises. The stern has poor visibility. The captain's position is blinded by the glare of the sun. That's what I am saying. It's my view, a matter about which I am knowledgeable and of which I am certain. When the wind is from that quarter, the bow is turned into the wind but the stern is turned directly to the glare of the sun. The captain's position has poor visibility. I however do not set the net on the outrigger side. And the thing I do is that I no longer align the canoe into the wind, but across the wind. That's the way to do it. But other questions are up to you and your own thinking. This is just a small thing I want to say briefly in reply to that captain.

(*Tokelau tautai* 31-32)

I commented earlier that I considered it incorrect to speak of contrasting VS and SV word-order patterns. Examination of the passages (14) and (15) shows that fronted NPs in verbal clauses are distributed as follows:

Transitive subjects	4
Intransitive subjects	3
Transitive objects	7
Obliques and genitives	6
Total	20

Transitive and intransitive subjects together account for seven, or less than half of fronted NPs. If we take the category absolutive, the unmarked case, we have 10, half the total number. So neither semantic nor grammatical subjects have a monopoly on this position. These figures demonstrate that in Tokelauan any divergence from verb-initial word order is not in the direction of exclusive subject-initial order. Moreover, irrespective of whether the employment of NP-initial order is increasing overall, I would maintain that the verb-initial, NP-initial opposition has long been present in Tokelauan in the discourse conditions described here, and will continue to operate in precisely this way.

Two points require comment. One is the occasional omission of ko before a fronted NP, for example (8.20), (9.7,14), and (11.4,5,23,24). In daily interaction in Tokelau, forms like Koe ka fano? *Are you going now?* are common. Even transcriptions of impromptu but solemn speeches by elderly men at formal occasions show quite frequent omission of ko. On the other hand, its psychological reality is attested by the habit of at least one native speaker, who will insert ko in a transcription when it is completely inaudible on tape.

Secondly, the status of clauses with preposed clitic pronoun subjects is uncertain. It is very unusual for such clauses to occur in a narrative sequence (there is an example at (19.2) below), but they are common in conversation and oratory, particularly with verbs of knowing and perception.

In spite of the frequency of NP-initial syntax in conversational and expository contexts in these languages, no-one so far as I know has raised the question of word-order change with respect to them. Presumably this is because ko is alive and well as a topic marker, and it is easy to demonstrate that ko-fronted clauses are syntactically derivative, from the simple 'basic' verb-initial type. In the two languages to which I now turn, Tuvalu and Tikopia, ko has lost its previous function as topic marker, so that NP-initial clauses lack any formal indication of the fact that they are anything other than basic sentence types.

Peter Ranby has commented (personal communication) on his surprise at the high proportion of NP-initial clauses, without benefit of ko, which he heard around him when he first arrived in Tuvalu in the early 1970s. In his *Nanumea syntax* (1973), example sentences which are verb initial somewhat outnumber those that are NP initial. (Possibly this reflects their origin in narrative or conversation respectively.) He distinguishes between fronted NPs marked with a for 'emphasis' (i.e. topicalisation), and those marked with ko for 'specifity' (i.e. focusing), a situation which corroborates Clark's analysis referred to above, page 269 (Ranby 1973:49). Similarly, example sentences in Besnier 1981, which appear from their content to be derived from conversational or elicitation contexts rather than narrative, are overwhelmingly NP initial. Besnier detected correlations between VSO order and formal registers (giving Bible translations as the most formal on a continuum), and SVO and informal registers, and suggested an ongoing change from VSO to SVO syntax.[2]

When we turn to narrative texts in Tuvaluan, a somewhat different picture is discernible. The texts consulted are of two kinds: Bible Society New Testament translations (*Maleko* 1969, *Ko te Tala 'Lei mō Aso Nei* 1977), and oral narratives collected a few years ago. I discuss these in turn.

The Bible translations show considerable variation in the use of ko as topic marker. In both versions we find most fronted NPs marked with the particle a. However ko is retained for topics after the conjunctions kafai *if*, *when* and ka *and*, *but*, and appears occasionally on other topicalised NPs. Focused NPs are marked with ko:

(16) A lesu ne afuli ne ia ki tua a te temoni gugu .... Ka ko nisi
I. T/A *cast out by him to outside the demon dumb but some*

0 latou ne fai ake penei: "ko te mana o Pelesepulo, te aliki o
of them T/A *say* dir. *thus the power of P. the chief of*

temoni, e afuli ei ne ia temoni mai i tino."
demons T/A *cast out* pro. *by him demons from in people*

Jesus was casting out a dumb demon .... But some of them spoke like this, "It is by the power of Beelzebub, the prince of demons, that he casts out demons from people."

(*Luke* 11.14-15)

No significant difference was noted in the number of fronted NPs in the 1969 edition of Mark and the 1977 edition of the New Testament. It is quite common however for a topic that is marked with ko in 1969 to be marked with a in 1977, for example te uaina fou (*Mark* 2.22).

The formality of the Bible translations is reflected in a less elliptical style, subject NPs being suppressed far less often than in oral narratives, and in the frequent use of the past tense marker ne on narrative clauses, alternating with the unmarked verb. A common pattern is for the first verb in a sequence of narrative clauses to be marked with ne, and for subsequent verbs to be unmarked. In either case, narrative clauses are almost always verb initial except when there is a strongly contrastive subject, as in (17.4) below. Background clauses, like those in (16), are frequently NP-initial.

(17) 1 Ne kati te ua o loane i te fale puipui,
T/A *sever the neck of I. in the prison*
John was beheaded in prison,

2 aumai ei tona ulu i te tisi,
bring pro. *his head on the dish*
his head was brought from there on a dish,

3 kae tuku atu ki te tamafine,
conj. *put* dir. *to the girl*
and given to the girl,

4 ka ko te tamafine tena ne tuku atu ne ia ki tona matua.
conj. top. *the girl that* T/A *put* dir. *by her to her mother*
and the girl took it to her mother.

5 Ne omai a soko o loane
T/A *come disciples of I.*
John's disciples came

6 o puke tona foitino
conj. *take his body*
and took his body

7 ke ave o tanu,
conj. *take* conj. *bury*
in order to take and bury it,

8 palele olo ei latou
then go pro. *they*
then they went

9 o taku atu ki a Iesu.
conj. *speak* dir. *to* *I.*
and reported it to Jesus.

10 | te logoga a Iesu i te tala i a Ioane,
on the hearing of I. *at the report about* *I.*
When Jesus heard the report about John's death,

11 fano ei a Iesu i te poti ki te koga lavaki,
go pro. *I.* *in the boat to the place desert*
he went by boat to a desert place

12 o nofo i ei tokotasi.
conj. *stay at* pro. *alone*
and stayed there on his own.

(*Matthew* 14.10-13)

Especially noteworthy in the Bible translations is the high proportion of NP-initial syntax in quoted direct speech, both in dialogue and in more extended passages such as the teachings of Jesus. The following examples are from the 1977 translation of *Mark:*

(18) A tino katoa konei e 'sala ki a koe
people all *here* T/A *search to* *you*
Everybody here is looking for you. (1.37)

Au e seki vau o kalaga ki tino amiotonu
I T/A *not come* conj. *call* *to people good behaviour*
I have not come to call just the righteous. (2.17)

A te Sapati ne fai mo te 'lei o te tino
the Sabbath T/A *make for the good of the person*
The Sabbath was made for the good of man. (2.27)

A tou matua mo ou taina kola e 'tu mai i tua
your mother and your brothers that T/A *stand* dir. *at outside*
Your mother and brothers are there waiting for you outside. (3.32)

A te tulafono tele e fakaaoga ne koutou o fakamasino ei nisi, e
the law *that* T/A *use* *by you* conj. *judge* pro. *some* T/A

toe fakaaoga ne te Atua o fakamasino ei koutou .... A te tino
again use *by the God* conj. *judge* pro. *you* *the man*

e isi ne ana mea ka toe fakaopoopo ana mea ke uke
T/A *be some his thing* T/A *again collect* *his things* conj. *plentiful*

The law that you use to judge others will be used in turn by God to judge you .... The man who has riches shall further increase his property. (4.24)

Turning to the two oral narratives, we find they bear a close resemblance to those of the languages considered earlier. As usual, the subject NP of a narrative clause will be suppressed if it is the same as the subject of the previous clause. Even switched subjects are suppressed if the referent is clear from the context. However, in narrative clauses which do have overt subjects, these invariably follow the unmarked verb, except in the (quite rare) case of clitic pronoun subjects. In the story of *Aoke*, there are 70 narrative clauses, of which 16 have expressed subjects, all post-verbal. In the story of *Kaimoko*, there are 60 narrative clauses, of which 23 have overt post-verbal subjects. So

between 23% and 34% of narrative clauses, even in the highly elliptical oral narratives, have post-verbal subjects. To assume that we cannot assign any constituent order type to a clause with suppressed NP constituents is unwarranted. Given the model of narrative structure outlined here, and the evidence of those narrative clauses which *do* have overt subjects, we are justified in assuming that all narrative clauses would be verb initial if there were no ellipsis. In the next example, notice that the first, fourth and ninth clauses have post-verbal subject NPs, whereas the direct speech at (19.16) is NP initial. Note also the absence of T/A marking on all except the background clause (19.12) and the direct speech clause.

(19) 1 Ffati ailoa i ai te tae a Kaimoko
break int. *at* pro. *the scoop-net K.*
Thereupon Kaimoko broke the handle of his scoop-net

2 kae na vvelo ngina
conj. *he spear* Cia
and he speared /him with it/

3 tuu ailoa i te mata
stand int. *at the point*
and stood erect on the point.

4 Takapili ake i ai uunaa a Lupo,
turn around dir. *at* pro. *there L.*
Lupo wheeled around there

5 makalo ailoa ki Tonga.
flash past int. *to T.*
and sped off to Tonga.

6 Fanatu
go dir.
He arrived there

7 ko te afanga ailoa o te tupu,
pred. *the canoe landing place* int. *of the king*
it was in fact the canoe bay of the king,

8 moe loa i ai.
sleep then at pro.
and he went to sleep there.

9 Taeao ake, kae fanaifo te aavanga a te tupu
morning dir. conj. *go* dir. *the wife of the king*
In the morning, the wife of the king came down,

10 o huahua telaa afanga.
conj. *rake their-2 beach*
and raked their canoe bay.

11 Huahua atu
rake dir.
She raked out,

12 e lave takitaki ailoa momea konaa
T/A *obstructed firmly* int. *places those*
but a certain place was firmly stuck,

13 kae pula atu
conj. *look* dir.
so she looked out

14 he tino
pred. *man*
/and saw it was/ a man.

15 Fanake ailoa ki uta i te tupu.
go dir. int. *to inland at the king*
She went inland to the king.

16 Ana muna: "Tupu! A te tino telaa e moe i ngaatai koa mate".
her words king the man that T/A *sleep at beach* T/A *dead*
Her words: "Oh King, that man who is sleeping on the beach is dead."

(*Kaimoko* 1-2)

Other examples of direct speech with NP-initial order from these two tales are:

(20) A koe ni taa ifea, i Samoa?
you T/A *kill whence from S.*
Where were you killed, Samoa?

Ko au e tuutuungina taula o vaka
top. *I* T/A *cut anchors of canoes*
I shall cut the anchor ropes of the canoes

Ko au ke fano o kau ilingina te matangi
top. *I* conj. *go* conj. *I fan the wind*
Let me go and cause the wind to blow

Taku taume, taku afi laa fanatu au o tutu
my spathe my fire int. *go* dir. *I* conj. *light*
My spathe, my fire, I went and set alight

The argument I am developing here can be summarised as follows. NP-initial syntax characterises the language of conversation, comment, exposition and other sorts of topic-oriented discourse. Verb-initial syntax characterises the language of narrative and other clauses in which the verb is emphasised, for antithesis or some other reason. The appearance of change from verb-initial to NP-initial syntax is to some extent an artifact of the type of data that has been examined at different times, and the kind of quantitative analysis that has been performed on it. Simple counting of word-order clause types is not adequate; the discourse type of the clause must also be taken into account.

Crucial support for this argument comes from Tikopia. In the case of this language we have two sets of data which are separated by a period of 50 years. The first set consists of oral narratives collected by Raymond Firth in 1928, and published in Firth 1961. The second set was collected in the late 1970s, and consists of oral texts recorded by Judith Macdonald in the Solomon Islands in 1979, and transcriptions of elicitation sessions conducted by linguists at Auckland University with Ishmael Tuki, a Tikopian schoolteacher, in 1978.

These data formed the corpus for Early's 1981 thesis. On the basis of a count of word-order patterns in simple independent clauses in this corpus, made without regard to the type of discourse being examined, Early concluded that Tikopia 'basic' word order is VS(O). However he expressed surprise at the predominance of NP-initial order in the Ishmael Tuki transcripts. He speculated

whether this might reflect a word-order change in Tikopia, resulting from increasing contact with Solomon Islands languages, or simply a tendency on the part of the informant to conform to the ordering of the language of his interlocutors, English.

The Tuki transcripts consist of isolated sentences produced in response to translation requests. In full declarative verbal sentences with expressed subject and/or object, not one proffered example is verb initial, except for four which have clitic pronoun subjects preceding the verb. Of the 56 NP-initial clauses, 75% are subject initial and 25% object initial. In my work with two Tikopian informants, in 1980, I also found NP-initial order to predominate. Out of 42 simple sentences elicited by me, only one had verb-initial order, and that was a case in which a clitic pronoun was suppressed. Early's suggestion of "foreigner talk" can be discounted. I discussed word order with one informant, and he was adamant in rejecting verb-initial versions of the sentences he supplied, except for yes/no questions:

(21)a. Te meleki ku maringi
the milk T/A *spill*
The milk has spilt

b. Ku maringi te meleki?
Has the milk spilt?

I mentioned earlier that the language of exposition and conversation is concerned with people, places, and things - topics in fact. The kinds of clause that linguists elicit from informants in the initial phases of a grammatical study are also of this type. They do not relate a sequence of events, but are statements or propositions about NPs. The following examples, taken from the Tuki transcripts, illustrate: *the man killed the pig, the man is dead, Bruce is not here, the chickens ate the coconut*, and so on. Such sentences will be interpreted by an informant as being 'about' the man, Bruce, and the chickens, and will tend to be encoded in NP-initial syntax.

The Firth texts consist of traditional narratives and, as expected, consist largely of verb-initial clauses. However they contain over 60 NP-initial verbal clauses. These do not occur in a random manner, but are distributed in the same way as in Tokelauan and Tuvaluan narrative. Two thirds occur in background clauses, and one third in direct speech. It is striking that almost every direct speech clause which is not imperative or exclamatory is NP initial:

(22) Na tana ke tangata ne fu ia.
now her own man T/A *hide* int.
Now, she had her man hidden indeed!

Koke au ki sea foki? Kuou e au ki se mana moku
you come to a what also I T/A *come to a thunder for me*
What are you too coming for? I have come for my thunder.

Marie, totou taina koi poi rei
wait your brother T/A *go then*
Wait, your brother is still going there.

Toku pokouru ku fai
my head T/A
My head has been split.

Au kai e nofo ke se ono ki ei?
your food T/A *sit you not look to it*
Your food is there, why don't you look after it?

(Firth 1961:185-194)

Of particular interest is a sequence of NP-initial dialogue clauses which foretell a series of happenings, and which are followed by a corresponding series of verb-initial narrative clauses which relate the occurrence of the same events (Firth 1961:193, lines 21-30).

The inescapable conclusion to be drawn from a careful study of the Firth texts is that NP-initial order was the normal or unmarked order for declarative clauses in conversation in 1928 - as it assuredly still is.

I now turn to the Macdonald texts. Since it could be argued that folk tales, and other traditional narratives which rely heavily on memorisation of whole sequences, might be expected to preserve archaic syntactic patterns, I shall not consider texts of this type from her collection. One of her texts is a spontaneous narration of personal experience, describing events of very recent occurrence and dramatic effect: "The story of Cyclone Kerry which hit Makira on February 15 and 16 1979". In this 340-word narrative there are 22 verb-initial narrative clauses. Even if we reject seven in which an agent NP is suppressed, we still have 15 verb-initial narrative clauses, as opposed to 13 NP-initial clauses, only three of which are of dubious status, the remainder being clearly background clauses containing orientating or evaluative content. So this text shows the same pattern of unmarked verb-initial narrative clauses, and T/A marked NP-initial background clauses, as the Firth texts of 1928.

(23) 1 Tā rei toku fasi fare i raro manioka
build then my piece house at under cassava
Then I built a half-house under the cassava.

2 Te rākau e tū, te voia e lasi,
the tree T/A *stand the v.* T/A *big*
There was a tree, a big voia tree,

3 kae oku makupuna e mēraki i roto paito
conj. *my grandchildren* T/A *sleep in inside house*
and my grandchildren were asleep inside the house.

4 Fati ko voia
break v.
The voia broke

5 ne pē ake fuarei fakatafa ma fasi paito
T/A *throw* dir. *just beside piece house*
it was just thrown down beside my half-house,

6 sakarukua ko a rā voia
break art. *branch v.*
some branches broke,

7 soroa rei ki vae paito
drop then to foot house
and dropped to the base of the house.

8 Sisene paia ko mātou
not hit we
We were not hit,

9 Ne sao fakatoaki fuarei ko mātou
T/A *safe narrowly just we*
we just narrowly escaped.

10 A ko mātou ka mate i te voia.
conj. *we* T/A *die at the v.*
But we could die of the voia tree.

11 Ta fafati mai ko te voia e matangi, na rua
then break dir. *the v. by wind its two*
Then the wind broke the voia tree, the second one.

12 Fenaifo
go dir.
It went down,

13 poi, tō rei pe nea sise ofā,
go fall then like thing not hurricaned
it went and fell then like something not struck by wind,

14 tō ifo arei ki na tafito voia
fall dir. *then to its base v.*
the voia fell down then to its base.

(*Cyclone Kerry* 1.14-20)

There is also an expository text in this collection, *The Youth Union of Nukukaisi*, which explains the aims and activities of a young people's work group. In this 800-word text the word order types are distributed as follows: ten verb-initial clauses (counting only those with expressed subjects), and 38 NP-initial clauses. There are also of course a number of nominal predicates which add to the total of NP-initial clauses.

It seems to me that this text, like the Tokelau fishing texts cited earlier, exemplifes an exceedingly flexible system in which the emphatic sentence-initial position can be assigned to any NP, or the verb, as the speaker's whim and the demands of the discourse dictate.

Clearly languages differ in the extent to which they avail themselves of the NP-initial option. I cannot imagine a Tokelauan informant rejecting out of hand verb-initial versions of declarative sentences, as my Tikopian informant did. What remains constant in the languages studied is the way the two clause types are distributed and the absolute connection of verb-initial syntax with narrative sequence. Samoan texts collected by Steubel in the late 19th century contain many NP-initial clauses. These always occur in the initial orientation sections of narratives, in other background clauses, in direct speech, and in exposition (for example *O le fa'alēmigao*, Steubel 1976:143).

I close with some remarks on Kapingamarangi and Nukuoro.[3] NP-initial clauses far outnumber verb-initial clauses in the Kapingamarangi narratives collected in Elbert 1948. Chung (1978:15-20) provides convincing arguments to show that verb-initial clauses are syntactically basic. Her statement that they are rare needs some qualification however. There are in fact many verb-initial clauses, but the majority of them have one or more NPs suppressed. One can of course assume that these clauses have 'underlying' initial subjects, but even in Kapingamarangi this may not be a necessary inference, when other facts are taken into account. A common pattern for the distribution of word-order types is as follows (the story *Ti kuru*, Elbert 1948:109, is a good example):

1) Clauses of which the subject is the same as that of the preceding clause are verb initial and the subject NP is suppressed.

2) Very rarely, such a clause will have an expressed, usually pronominal, subject, which follows the verb.

3) The subjects of subject-initial clauses are new subjects, i.e. not the same as the subject of the preceding clause, and thus suitable targets for topicalisation.

4) Occasionally, as Chung notes, other types of NP occur pre-verbally, leaving the subject in post-verbal position.

5) Sometimes verb-initial clauses are clearly the result of a need to emphasise the verb.

The data are consistent with a system in which sentence-initial position is reserved for topicalised constituents, and in which new subjects are favoured for topicalisation. Such a system appears to be an intermediate step between that prevailing in the other languages I have considered, and that of Nukuoro, which on the basis of the one text available to me (Carroll 1965) appears to be a wholesale convert to NP-initial order. Clauses with non-expressed subjects are comparatively uncommon, pronominal subjects being repeated in pre-verbal position in successive narrative clauses. There are however some interesting exceptions to this pattern:

1) NPs in other syntactic functions sometimes occur clause initially.

2) Certain transitive verbs (all the examples are of perception verbs) seem to require a post-verbal pronoun copy of the sentence-initial subject NP.

3) Of particular interest is the use of a post-verbal particle hu in certain verb-initial clauses. This seems to confer verb focus:

(24) Gai Iaidemalo ga hai ange bolo gilau gi aahe. Gai ga aahe
then I. T/A *say* dir. *that they*-2 opt. *turn back* *then* T/A *turn back*

adu hu gilau.
dir. *they*-2

Then Iaidemalo said that they should turn back. And turn back they did.

(Carroll 1965:459)

There is nothing in the Nukuoro narrative to contradict the proposal that NP-initial order is topic-initial order. Note that indefinite subjects such as those of negative existential clauses, which are unsuitable candidates for topicalisation, always occur post-verbally. The use of the particle hu to highlight a clause-initial verb looks like an intriguing reversal of the situation in other languages in which ko is used to highlight a clause-initial noun phrase.

SOURCES

I wish to thank the following people for making available their unpublished transcriptions of oral texts: Bruce Biggs (Futunan), Keith and Anne Chambers (Tuvaluan), Antony Hooper and Logotasi Iosefa (Tokelauan), and Judith Macdonald (Tikopian). Unpublished sources cited in the paper are listed below, together with the name of the story-teller and the date of recording.

Tokelauan: *Tioita* and *Hola* (Pulu Levao, February 1984). *Tala faiva* (Ropati Simona, August 1983). *Tokelau fishing* and *Tokelau tautai* (Peato Perez, August 1977).

Futunan: *Talatuku o sigave* (Manafa Malau Lafaele, 1981). The cave of Lita. An extract from the East Uvean newspaper, *Fetuuaho*, October 13, 1973.

Tuvaluan (Nanumea): *The Story of Aoke and Metometo* (Fiapule 1974). *Kaimoko and Lupo and the Tongan Raid* (Vailei 1974).

Tikopia: *The story of Cyclone Kerry* (Pae Rongomao, 1979). *The Youth Union of Nukukaisi* (Paul Firiausi, 1979).

The glosses contain the following abbreviations for grammatical terms:

art.	article	opt.	optative
cav.	caveat	pl.	plural
conj.	conjunction	pred.	predicate marker
dir.	directional particle	pro.	proform
foc.	focus	Q	question introducer
int.	intensive particle	T/A	tense-aspect particle
neg.imp.	negative imperative	top.	topic

NOTES

1. I am grateful to Ross Clark, Keith Chambers, Peter Ranby, Andrew Pawley, Frank Lichtenberk and Niko Besnier for helpful comments on this paper.

2. This position has since been revised in favour of the view that word-order variation in Tuvaluan is governed by pragmatic and semantic factors. See Besnier 1986, section 6.

3. I leave out of consideration certain Outlier languages in which the consistent use of SVO order is usually considered to be the result of influence from neighbouring non-Polynesian languages: Emae, Mele-Fila, and Futuna-Aniwa (Clark 1986).

REFERENCES

BESNIER, Niko

1981 Transitivity and case in Tuvaluan. M.A. thesis, Stanford University.

1986 Word order in Tuvaluan. In this volume, 245-268.

CARROLL, Vern

1965 An outline of the structure of the language of Nukuoro. *Journal of the Polynesian Society* 74:192-226, 451-472.

CHUNG, Sandra

1978 *Case marking and grammatical relations in Polynesian.* Austin: University of Texas Press.

CLARK, Ross

1976 *Aspects of Proto-Polynesian syntax.* Auckland: Linguistic Society of New Zealand.

1986 Linguistic convergence in Central Vanuatu. In *FOCAL II,* 333-342.

EARLY, Robert J.

1981 Minor morphemes of the language of Tikopia. M.A. thesis, University of Auckland.

ELBERT, Samuel H.

1948 *Grammar and comparative study of the language of Kapingamarangi, texts and word lists.* Washington: Pacific Science Board, National Research Council.

ELBERT, Samuel H. and Torben MONBERG

1965 *From the two canoes.* Honolulu and Copenhagen: University of Hawaii Press and Danish National Museum.

FIRTH, Raymond

1961 *History and traditions of Tikopia.* Wellington: The Polynesian Society, Memoir No.33.

FORSYTH, J.

1970 *A grammar of aspect: usage and meaning in the Russian verb.* Cambridge: The University Press.

HOPPER, Paul J.

1979a Aspect and foregrounding in discourse. In T. Givón, ed. *Syntax and semantics,* 12: *Discourse and syntax,* 213-241. New York: Academic Press.

1979b Some observations on the typology of focus and aspect in narrative language. *Studies in Language* 3:37-64.

1982 Aspect between discourse and grammar. In Hopper, ed. 1982:3-18.

HOPPER, Paul J., ed.

1982 *Tense-aspect: between semantics and pragmatics.* Amsterdam and Philadelphia: John Benjamins.

HUNTSMAN, Judith

1977 *Ten Tokelau tales*. University of Auckland, Department of Anthropology Working Papers No.47.

LABOV, William

1972 The transformation of experience in narrative syntax. In his *Language in the inner city*, 354-396. Philadelphia: University of Pennsylvania Press.

NEW TESTAMENT

1977 *Ko te Tala 'Lei mō aso nei*: The New Testament in Tuvalu. Suva: The Bible Society in the South Pacific.

RANBY, Peter

1973 Nanumea syntax. M.A. research essay, University of Auckland.

SCHIFFRIN, Deborah

1981 Tense variation in narrative. *Language* 57:45-62.

SILVA-CORVALÁN, Carmen

1983 Tense and aspect in oral Spanish narrative: context and meaning. *Language* 59:760-780.

STEUBEL, C.

1976 *Myths and legends of Samoa*. Wellington: A.H. and A.W. Reed.

WOLFSON, Nessa

1979 The conversational historical present alternation. *Language* 55: 168-182.

THE ACTOR EMPHATIC CONSTRUCTION OF THE EASTERN POLYNESIAN LANGUAGES

Ray Harlow

1. Among the features which set the Eastern Polynesian (EP)[1] languages apart as a subgroup of the Polynesian (PN) language family is the productive use of a peculiar construction usually referred to as Actor Emphatic.[2] The general shape of this construction and the differences between it and unmarked sentence structures can be illustrated by means of these examples from Maori:

(1) I tiihore a Pita i te hipi.
T *skin* art. *Peter* obj. art. *sheep*
Peter skinned the sheep.

(2a) Na Pita i tiihore te hipi.
Prep. *Peter* T *skin* art. *sheep*
It was Peter who skinned the sheep.

(2b) Na Pita te hipi i tiihore.
ditto

Similarly,

(3a) Ma Pita e tiihore te hipi.
Prep. *Peter* T *skin* art. *sheep*
Peter will skin the sheep. or *Peter is to skin the sheep.*

(3b) Ma Pita te hipi e tiihore.
ditto (examples from Clark 1976:111)

Sentences (2a and b) and (3a and b) are examples of Actor Emphatic. As illustrated in sentence (1), the unmarked verbal sentence patterns in MAO (and similarly in the other EP languages) involve:

(i) VSO order,
(ii) unmarked subject, and
(iii) an object marker i in transitive clauses.

Actor Emphatic sentences differ from these in that:

(i) the order is SVO or SOV,
(ii) the subject is marked with a preposition, and
(iii) there is no object marker.

Obviously, this formulation begs a question or two, in particular, what exactly are the grammatical relations in this type of sentence,[3] and are sentences

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics,* 297-308. *Pacific Linguistics*, C-93, 1986.

of this type derived from corresponding unmarked active (or passive)[4] sentences in the Transformational sense; I shall refer to the NPs in these constructions as NP_1 and NP_2. Thus, the pattern for Actor Emphatic in MAO is:

(4) Na NP_1 i V NP_2.
NP_2 i V.

Ma NP_1 e V NP_2.
NP_2 e V.

Further features of Actor Emphatic in MAO not revealed by these examples are the restriction of the tense-aspect markers to i and e, whereas in sentences of the type (1), there is a considerably more extensive paradigm available,[5] the corresponding semantic restriction to past or future-cum-modal uses and the restriction of the construction to transitive verbs (though see below). The prepositions used to mark NP_1 have other uses as well, primarily in predicative possessive constructions, e.g. (examples again from Clark 1976:112-113):

(5) Na Pita te pukapuka.
Prep. *Peter* art. *book*
The book belongs to Peter.

(6) Ma Pita te pukapuka.
Prep. *Peter* art. *book*
The book is for Peter.

The comparable constructions in the other EP languages agree with the MAO construction in having NP_1 in initial position marked with a preposition or prepositions otherwise used for predicative possessive constructions;[6] in other details, however, a certain amount of variation is evident.

In this paper, I want to hazard a reconstruction of the PEP innovation reflected in the Actor Emphatic sentence types of the EP languages, to give some account of the developments in the individual EP languages, and to speculate about the origin of the innovation. In doing so, I shall not be departing radically from the conclusions reached by Clark (1976), which is so far as I know the only previous treatment of this construction across several languages. Like Clark, I have relied almost entirely on published grammars and texts for the languages I looked at. More rigorous work with informants may well lead to amendments to some of the conclusions reached, but, I suspect, not to radical changes.

2. Allowing that the reconstruction of aspects of the syntax of non-attested proto-languages is fraught with controversy, being regarded by some as impossible in principle, and has not got well established techniques such as those associated with the reconstruction of phonological and morphological systems, it does not seem unreasonable to attribute to the proto-language at least those features on which all daughter languages agree. On this basis, sentences of the shapes (7) and (8) were grammatical in PEP:

(7) 'a NP_1 i V NP_2.

(8) Ma'a NP_1 e V NP_2.

where NP_1 is [+human], V a canonical transitive verb, and NP_2 definite.

I want further to claim that at least at the time of the innovation which led to the existence of such sentences and probably in PEP itself these were the only such sentences. This claim is advanced on two unrelated grounds; firstly, these are the only patterns on which all EP languages surveyed agree, and it seems to me that the various departures from this pattern exemplified in the daughter languages (and we shall see that there are six types of such departure) are easily accounted for in terms of a reanalysis in at least some EP languages and plausible analogical extensions in all EP languages. The converse position, that the construction was less constrained in one or the other of these six ways in PEP and underwent restrictions in its development in one or more daughter languages, appears harder to justify. Secondly, the parallels I shall adduce below (section 4) seem to suggest a natural connection between possessive constructions and high transitivity in the sense of Hopper and Thompson (1980), and thus render more plausible this reconstruction of the initial innovation than one of lower transitivity.

Clark (1976:119) suggests a constituent structure for these sentences, taking the MAO sentence (2a) as an example, along these lines:

(9) $[_{S}[_{PRED}$[Na Pita][i tiihore]]$_{PRED}[_{NP}$te hipi$]_{NP}]_{S}$

The exact structure of the Predicate remains obscure, in particular, whether the verbal expression i tiihore is a relative clause on Pita or some sort of non-finite, quasi-participial thing. However, the main thing here is the claim that NP_2 is subject of some sort of extended possessive predicate. Again, this view is supported by the two arguments mentioned above. Firstly, for those languages which clearly treat NP_2 as an object, a relatively straightforward reanalysis can be postulated to explain this (cf. Clark 1976:121), whereas the converse is not so. And secondly, the innovation generally, especially the use of the possessive prepositions, is intelligible only if Actor Emphatic arose initially from a predicative possessive construction.

3. As indicated above, all the EP languages surveyed exhibit departures from the patterns postulated as the original form of the innovation. In all cases, the original patterns remain grammatical, and the departures represent relaxations of restrictions on these patterns so that other types are grammatical as well. These developments seem to be of two types, each of which is readily understandable, though the intersection of these two types in any one language leads to a squishiness of grammatical relations, to which I shall return below.

The two types are:

(i) The grammatical relations and sentence structure remain the same as in the original pattern, but selectional restrictions on NP_1, and the restrictions of the tense-aspect markers to i and e and the aspect to telic are relaxed, and NP_2 may be indefinite. That is, the availability of the construction in its original shape is extended to sentences of lower transitivity. Examples of the first such development can be found in RAR and MVA, where nominalisations occur as NP_1, and in MAO and RAR, where non-human and even non-animate nouns occur in this position, and the role of NP_1 is clearly no longer purely agentive, but is extended to something like cause or force in the Case Grammar sense. Thus:

(10) Na te 'akairo o te reta-topa (') e 'akapāpū mai ki a
Prep. art. *write* *of* art. *glottal stop* T *make certain hither to* art.
tātou i te tika'anga tikāi i te 'akatangi'anga o te reira
us obj. art. *correctness just* obj. art. *pronunciation of* art. *there*
tuatua. (RAR Simiona n.d.:iv)
word
The writing of the glottal stop (') will make certain for us the correct pronunciation of that word.

(11) Na te ua rāua i 'akaara. (RAR Buse 1963b:401)
Prep. art. *rain they2* T *awaken*
It was the rain that woke them.

Tense-aspect markers other than i and e, usually e-ana and e-nei, and usually in an habitual or imperfective sense, are possible in MVA, HAW, RAR, TAH, TUA and MRA, e.g.:

(12) Na'ai e 'aka'oro ana i te rauiti. (RAR Rere 1961:49)
Prep.-*who* T *drive* T obj. art. *launch*
Who drives the launch?

And finally, in MAO, NP_2 may be indefinite. Thus:

(13) Maa-ku e koorero he tikanga. (Chung 1978:179)
Prep.-*me* T *speak* art. *advice*
I will tell (you) a piece of advice.

This type of development is consistent with the highly subject-like behaviour of NP_2 in MAO,[7] and the occurrence of an alternative order: NP_2 T V in HAW, MAO, RAR and TAH. Chung (1978:179), in discussing MAO, calls this Raising from an underlying structure of the form:

(14) $[_S[_{PRED}\text{Prep. } NP_1]_{PRED}[_S T\ V\ NP_2]_S]_S$

but Clark (1976:119), rightly, connects this alternative order to the more general phenomenon, "extraposition of the second constituent of a complex predicate over a short subject".

(ii) The second type of development is a reanalysis along the lines suggested by Clark (1976:121f). Clark proposes that in at least some EP languages a reanalysis has occurred such that from an unmarked active transitive sentence like the MAO one illustrated as (1), Actor Emphatic sentences are derived by a rule fronting the subject and marking it with the appropriate possessive preposition, leading to a structure like this:

(15) $[_S[_{PRED}\text{Na Pita}]_{PRED}[_S[_{PRED}\text{i tiihore}]_{PRED}[_{PP}\text{i te hipi}]_{PP}]_S]_S$.

Now, Actor Emphatic sentences like this, i.e. with NP_2 marked by i 'obj.' are not grammatical in MAO, but are in all the other EP languages surveyed except MQA. Indeed, in most of them this is the preferred pattern, though note that in the order NP_2 T V, if it occurs at all, NP_2 is never marked 'obj.'. Further, many of these languages have Actor Emphatic-like sentences involving intransitive

verbs like *go, speak*, whose subjects are agents. MVA has perhaps progressed furthest in this direction in allowing Actor Emphatic with verbs like pure *pray*, meimata *weep*, but similar sentences are possible in apparently all but MQA, TUA, and MRA. An interesting example from TUA (Stimson 1933a:35):

(16) Na te tahuga anake e tika kia korero i te igoa tapu o Kio.
Prep. art. *priest only* T *right* comp. *speak* obj. art. *name holy of Kio*
Only the priest may pronounce the holy name of Kio.,

shows that initial position and marking with na may be no more than a device for focussing agentive subject phrases, and have little more to do with the original construction.

I want to propose that what I have been calling two different types of development here are in fact the result of a single change in progress, namely the reanalysis, and that the development of Actor Emphatic in the EP languages is another example of gradual syntactic change of the sort discussed by Chung (1978:319) for Pukapukan, a Samoic PN language. During the course of such changes, in which a reanalysis occurs involving shifts in grammatical relations, the NPs concerned can show behaviour characteristic of both the old and the new grammatical relations. In the case of Pukapukan, it is contended, the reanalysis is of Passive as Ergative, in our case here, of an extended predicative possessive sentence with a patient NP as subject, as a transitive (or even intransitive) pattern with a fronted subject and, when present, the patient as object. The possibility that one is dealing with change in progress, at least in MAO, is suggested by the variability of native speaker judgements over sentence types at the forefront of the change. Thus, while there is no quarrel with sentences like (2a and b) and (3a and b) above, with sentences involving intransitive verbs, indefinite NP_2s, or even occasionally with i 'obj.', opinions differ strongly.[8]

Clearly, some languages have progressed further in this direction than others. MAO and MQA seem to be the most conservative (though see below for the possibility of a different reanalysis in MQA), whereas in others, such as TUA, TAH and RAR, the features reflecting the proposed original pattern are rather fewer. Striking is the fact that those languages which have expanded the tense-aspect paradigm for Actor Emphatic sentences to include markers used in unmarked main clauses in habitual or progressive senses are exactly those where NP_2 is more frequently marked with i 'obj.' than unmarked.[9]

Given this account, a question arises as to the time of the beginning of the reanalysis or better, of the development which may culminate in a complete reanalysis of Actor Emphatic patterns. One is I think faced with two equally unattractive possibilities; either it began in PEP, in which case it has been in progress for quite some time in languages like MAO, or it is a spontaneous independent development in at least a number of the daughter languages of PEP. Perhaps in favour of the second scenario is the fact that in MQA, a different reanalysis may be taking place. In a number of examples of Actor Emphatic given in Dordillon (1931), the verb is marked with the passive suffix, e.g. (p.66):

(17) Na te Etua i pepena tia te áni.
Prep. art. *God* T *create* pass. art. *heaven*
Le ciel a été créé par Dieu.

cf. also Lavondès (1966,vol.1:35):

(18) 'u ke'ahi 'ia na to matou tuakana.
T *kick* pass. prep. art.-*of us* *older* SSS
nos soeurs aînées l'ont frappé à coups de pieds.

From such examples, it would appear possible that in MQA the original pattern has been reanalysed as derived from an unmarked passive with a (not even always) fronted agent marked with na. Not wanting to go into this further, I can do no more here than point to the similarity of this state of affairs to the well-known 'drift' of several Indo-European languages from synthetic to analytic.

4. It may ultimately be futile to speculate about the origin of the Actor Emphatic construction, but it is interesting to do so. Clark (1976:119-121) sketches a proposal, but is clearly unhappy with it. This consists in asserting that the Actor Emphatic construction is a simple predicative possessive construction, such as is found in all PN languages, with some sort of clausal elaboration related to the possessive-relative. In all PN languages a genitive with a clausal expansion of which the NP in the genitive can be understood as the subject can be used as a relative clause within a NP, and even as a headless relative clause, as:

(19) ko e haa ho'o me'a 'oku kai. (Tongan, Clark 1976:118)
Topic art. *what your thing* T *eat*
What are you eating?

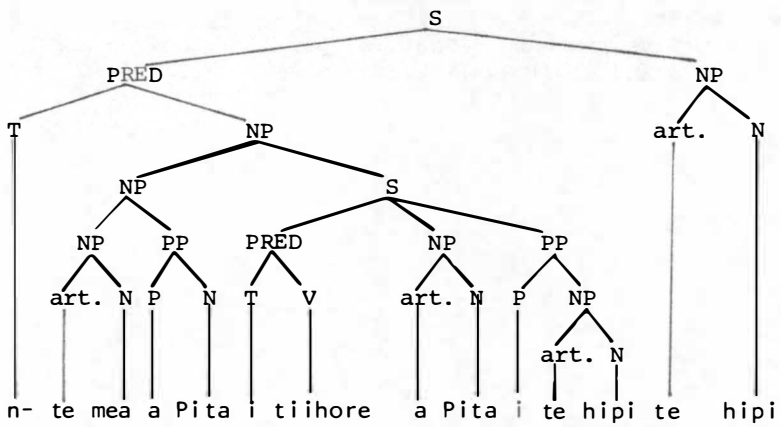
(20) Ko eenei ngaa riiwai a Hata i kawe atu ai i te paakoro ki te
Topic *these* art. *potato of H* T *carry away* Pro. *from* art. *shed* *to* art.
rori. (MAO, Bauer n.d.:19)
road
These are the potatoes which Hata carried from the shed to the road.

However, Clark's analysis of these as coming from two independent relative clauses, one of which becomes the genitive, and the other of which becomes the clausal part by deletion of the coreferential subject, is unable to account for the superficially similar predicative use in the Actor Emphatic constructions of the EP languages. He attempts to remedy this inability of his analysis of genitives to account for Actor Emphatic by proposing that predicative genitives (from which his attributive ones are derived) are derived in turn from something like:

(21) $[_{PRED}\ T[_{NP}[_{NP}\text{art. N}]_{NP}[_{PP}\text{a/o NP}]_{PP}]_{NP}]_{PRED}$

where art. is deleted following a T and the N is semantically empty. According to this view, Actor Emphatic is derived from a structure like:

(22)



Apart from the misgivings expressed by Clark himself, there are two reasons why neither an historical nor a synchronic derivation of Actor Emphatic from such a structure is satisfactory. Firstly, this possessive-relative clause is not restricted to transitive objects, but may be used to relativise on a variety of NP relations which cannot be NP_2 in an Actor Emphatic sentence, e.g.

(23) Te 'aso o Mautikitiki noko hano ai ki tai. (Rennellese, Clark 1976:117)
art. *day of M* T *go* Pro. *to sea*
The day that Mautikitiki went to sea.

(24) Te vāhi tā'u e haere atu. (TAH, Coppenrath and Prévost 1974:272)
art. *place* art.-*of-me* T *go away*
l'endroit où j'irai.

Secondly, in many languages (though not all) a pronominal copy ai of the NP relativised on is present, something never found in Actor Emphatic; e.g. sentences (20) and (23) above.

It seems to me that despite their superficial similarity, the origins of the Actor Emphatic construction should not be sought in the possessive-relatives, but rather in another quarter, a phenomenon which can be frequently observed in a variety of languages, the relationship between possessivity and transitivity. It is striking how often transitive, especially perfective transitive, constructions involve auxiliaries, case markings, personal affixes, etc., which are otherwise used in possessive constructions. It suffices here to allude to the *have* perfects in Romance and Germanic languages and in Greek, which arguably developed independently, and the work of people like Allen (1964), Anderson (1977), Seiler (1983),[10] and other work cited there, which shows that this phenomenon is by no means restricted to these languages. A common feature of this type of thing is the formal parallelism of agent and possessor on the one hand and patient and possessum on the other. To my knowledge, no-one has tried to account for this, and I shan't try to myself, except to point out that these roles are at least extensionally very similar, in that those entities which are typically agents are exactly those which are typically possessors, and similarly for patients and possessa.

All I want to do here is to point to the parallelism of the Actor Emphatic construction in EP languages and the phenomenon I have mentioned and thus to explain Actor Emphatic as arising through some sort of universal possibility of

extending sentential possessive constructions with a verbal expression to form a (usually) perfective transitive sentence. That is, parallel to (25a and b) and (26a and b), there is (27a and b) (= (2a)).

(25a) nora tun ē
he-gen. *house is*
He has a house.

(25b) nora bereal ē
he-gen. *carry* non-fin. *is*
He has carried (something).

but

(25c) sa ekeal ē
he-nom. *come* non-fin. *is*
He has come. (Armenian)[11]

(26a) Mihi est filius / Filium habeo
Me-dat. *is son*-nom. / *son*-acc. *I have.*
I have a son.

(26b) Mihi est res tota provisa / Rem totam provisam habeo.[12]
Me-dat. *is thing*-nom. *whole provided / thing*-acc. *whole provided I have*
I have provided for the whole matter. (Latin)

(27a) Na Pita te hipi
Prep. *P* art. *sheep*
The sheep belongs to Peter.

(27b) Na Pita i tiihore te hipi (=2a) (MAO)

Against this as an 'explanation', there are no doubt several possible objections; for instance that it does not solve the question of the status of the VP in Actor Emphatic sentences. One possible line of attack, however, I do want to try to answer, and that is the point that the parallels adduced from Latin, Armenian, etc. are all translations of English *have*, whereas the PN possessive type of which Actor Emphatic is an extension is not, but rather a translation of English *belong to*. Oddly, constructions which have one of these uses in one language sometimes have the other in other languages, e.g. *NP_1-nom. is NP_2-gen.* in Armenian = 'NP_2 has a NP_1', but in Latin = 'NP_1 is NP_2's' or 'NP_1 belongs to NP_2'. Similarly, *NP_1-nom. is NP_2-dat.* in French or Swiss German = 'NP_1 belongs to NP_2', but in Latin = 'NP_2 has a NP_1'. I am aware that this is more a curiosity than a convincing proof, but want nonetheless to claim that the parallelism between the postulated origin of Actor Emphatic and the use of *have*-like possessives as transitives in other languages can be upheld, because the sole difference between *have a* sentences and *belong to* sentences is not a matter of the predicate but only of the topicality and rhematicity of the arguments. That is, that the only difference between the uses of sentences like Latin (26a) and MAO (27a) is that the Latin sentence is 'about' me, and asserts the owning of a son, while the MAO sentence is 'about' the sheep and asserts its being owned by Peter. If that is so, the claim that Actor Emphatic is like the *have* transitive perfects and related phenomena in a variety of languages can stand.

Clearly, even if right, this paper accounts for Actor Emphatic only by assigning it to a group of phenomena which need explaining as a whole. That this doesn't seem to have happened yet does not suffice to dispel the impression that there is something in the way of a widespread tendency or even universal possibility involved here.

NOTES

1. On the subgrouping of the PN languages, see Pawley 1966 and 1967, and Green 1966. The EP languages used in this study, along with abbreviations and sources are:
   Maori: MAO: Chung 1978, Bauer 1981 and n.d., Karaka Roberts, personal communication.
   Rarotongan: RAR: Buse 1963a,b and c, Rere 1961, Simiona n.d.
   Tahitian: TAH: Coppenrath and Prévost 1974.
   Hawaiian: HAW: Elbert and Pukui 1979.
   Marquesan: MQA: Lavondès 1966, Dordillon 1931.
   Mangarevan: MVA: Janeau 1908.
   Manihiki: MRA: Kauraka 1982.
   Tuamotuan: TUA: Stimson 1933a and b.
   Easter Island: EAS: Mulloy and Rapu 1977, Chapin 1978, Fuentes 1960.
   All of these apart from EAS are believed to belong to a subgroup of EP called Central Eastern Polynesian (CEP), which in turn divides into Tahitic (TA), consisting of MAO, MRA, RAR, TAH, TUA, and Marquesic (MQ), consisting of HAW, MVA, MQA.
   In the glosses of examples, the following abbreviations are used:
   T = tense-aspect marker
   art. = article, proper or common
   obj. = object marker
   prep. = preposition other than obj.
   pass. = passive suffix
   comp. = complementiser
   Pro. = pronominal copy.

2. Two examples of superficially similar constructions in non-EP languages are known: Luangiua maa ke poi i ŋamuŋamu ke ivi *It is for the dog to chew the bone.* quoted in Clark 1976:123, and Tikopia aaku ne taa *I hit it* in Early 1981:97. There is at present simply not enough information to tell whether these are at all related to the EP construction under discussion here.

3. On this question, see Chung 1978:177ff and Bauer 1981:321ff.

4. cf. Hohepa 1967:102 and Buse 1963c:637.

5. As Chung 1978:177 points out, e as T occurs only in embedded clauses and is not available as T in main clauses. On MAO tense-aspect markers, cf. Bauer 1981:52ff.

6. The precise form of the prepositions varies. In EAS, the pair is 'a͡uma, continuing the forms in '- and m- of PEP. PCE however innovated by introducing n- giving the pair na͡uma as in MAO. This pair is preserved only in MAO, MVA and Penrhyn, merging in all other EP languages to na, both in Actor Emphatic constructions and in predicative possessive uses. Similarly, though without relevance here, with the o-class prepositions.

7. See esp. Chung 1978:177ff.

8. e.g. Chung 1978:180 quotes one example where NP_2 is marked with i and is reflexive, and (p.181) another in which NP_2 is marked with ki and is 'object' of titiro *look*, arguably not a canonical transitive, but a 'middle' verb, cf. Chung 1978:478. However, my informant rejects both of these in favour of unmarked NP_2s, but allows the very similar: Maa koutou e whakarongo mai ki a au *You should listen to me.*

9. Apart from one possible example of Actor Emphatic where T = ka 'inceptive' in MVA, none of the other tense-aspect markers available in unmarked sentence types are attested.

10. On Greek, cf. also Aerts 1965.

11. Evidently, Armenian has undergone a reanalysis similar to that posited for EP languages, cf. Anderson 1977:340; the genitive form has been reanalysed as transitive subject and the patient of such sentences stands in the accusative.

12. The first sentence in (26b) is from Cicero *Verr.* IV 42, 91. The second is an unattested but grammatical Vulgar Latin sentence.

REFERENCES

AERTS, W.J.

1965 *Periphrastica. An investigation into the use of* εἶναι *and* ἔχειν *as auxiliaries or pseudo-auxiliaries in Greek from Homer up to the present day.* Amsterdam: Hakkert.

ALLEN, W.S.

1964 Transitivity and possession. *Language* 40:337-343.

ANDERSON, S.R.

1977 On mechanisms by which languages become ergative. In C.N. Li, ed. *Mechanisms of syntactic change*, 317-363. Austin: University of Texas Press.

BAUER, W.

1981 Aspects of the grammar of Maori. Ph.D. thesis, University of Edinburgh.

BUSE, J.W.

1963a Structure of the Rarotongan verbal piece. *BSOAS* 26:152-169.

1963b Structure of the Rarotongan nominal, negative and conjunctival pieces. *BSOAS* 25:393-419.

1963c Rarotongan sentence structure. *BSOAS* 26:633-645.

CHAPIN, P.G.

1978 Easter Island: a characteristic VSO language. In W.P. Lehmann, ed. *Syntactic typology: studies in the phenomenology of language*, 139-168. Hassocks, Sussex: Harvester Press.

CHUNG, S.

1978 *Case marking and grammatical relations in Polynesian.* Austin: University of Texas Press.

CLARK, R.

1976 *Aspects of Proto-Polynesian syntax.* Auckland: Linguistic Society of New Zealand.

COPPENRATH, H. and P. PRÉVOST

1974 *Grammaire approfondie de la langue tahitienne (ancienne et moderne)*. Papeete: Librairie Pureora.

DORDILLON, R.I.

1931 *Grammaire et dictionnaire de la langue des îles Marquises*. Travaux et Mémoires 17. Paris: Institut d'Ethnologie.

EARLY, R.J.

1981 Minor morphemes in the language of Tikopia. M.A. thesis, University of Auckland.

ELBERT, S.H. and M.K. PUKUI

1979 *Hawaiian grammar*. Honolulu: University Press of Hawaii.

FUENTES, J.

1960 *Dictionary and grammar of the Easter Island language*. Santiago de Chile: Editorial Andres Bello.

GREEN, R.C.

1966 Linguistic subgrouping within Polynesia: the implications for prehistoric settlement. *Journal of the Polynesian Society* 75:6-38.

HOHEPA, P.W.

1967 *A profile-generative grammar of Maori*. Indiana University Publications in Anthropology and Linguistics, Memoir 20. Bloomington.

HOPPER, P.J. and S.A. THOMPSON

1980 Transitivity in grammar and discourse. *Language* 56:251-299.

JANEAU, V.F.

1908 *Essai de grammaire de la langue des îles Gambier ou Mangaréva*. Braine-le-Comte, Belgium: Zech et fils.

KAURAKA, K.

1982 *Tales of Manihiki*. Suva: Institute of Pacific Studies.

LAVONDÈS, H.

1966 *Récits marquisiens (2ème série)*. Papeete: Office de la Recherche Scientifique et Technique Outre-Mer.

MULLOY, E.R. and S.A. RAPU

1977 Possession, dependence and responsibility in the Rapanui language. *JPS* 86:7-25.

PAWLEY, A.

1966 Polynesian languages: a subgrouping based on shared innovations in morphology. *JPS* 75:39-64.

1976 The relationships of Polynesian Outlier languages. *JPS* 76:259-296.

RERE, T.

1961 *Conversational Maori, Rarotonga language*. Rarotonga: Cook Islands Government.

SEILER, H.-J.

1983 Possessivity, subject and object. *Studies in Language* 7:89-117.

SIMIONA, T.

n.d. *E au tua ta'ito nō te Kūki 'Airani*. Suva: Institute of Pacific Studies, University of the South Pacific.

STIMSON, I.F.

1933a *Tuamotuan religion*. Bernice P. Bishop Museum Bulletin 103. Honolulu.

1933b *The cult of Kiho-tumu*. Bernice P. Bishop Museum Bulletin 111. Honolulu.

Harlow, R. "The actor emphatic construction of the Eastern Polynesian languages". In Geraghty, P., Carrington, L. and Wurm, S.A. editors, *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*. **C-93:297-308**. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.297*

INTELLIGIBILITY PATTERNS IN SABAH AND THE PROBLEM OF PREDICTION

Paul R. Kroeger

1. INTRODUCTION

The language or dialect boundaries which exist in a given geographical area are not necessarily a barrier to communication among the language groups of the area. What factors enable one speech group to understand the language of a neighbouring group? What factors determine the extent to which a group will be understood by its neighbours?

Intelligibility or comprehension across linguistic boundaries is a very complex phenomenon. Linguistic similarity, social contact, language attitudes, patterns of language use, educational policies and political pressures are some of the factors relevant to explaining intelligibility. Casad (1974), Collier (1977), Simons (1979) and others have been interested in developing models for predicting intelligibility from various linguistic and sociolinguistic measurements. Linguistic surveys, which may be conducted for many reasons, are the source both of the initial data for developing such models and of further data for testing them.

Statistical analysis of intelligibility testing data is a natural approach to the development of empirical models. However, the very nature of a dialect intelligibility survey places constraints on the data collected which call into question the applicability of some statistical procedures.

In this paper I will discuss the results of one particular linguistic survey, using various kinds of statistical measurements which help us interpret the data. But problems were encountered in applying some of the more sophisticated statistical procedures to intelligibility scores. The use of regression analysis with intelligibility data, particularly when the results are generalised beyond a particular sample, seems to be especially prone to error.

In most contexts, I am using the word *intelligibility* simply as a synonym for comprehension. At times it is necessary to distinguish *inherent intelligibility*, defined by Simons (1979) as "the theoretical degree of understanding between dialects whose speakers have had no contact", from learned comprehension due to language contact. At such times, the term *bilingualism* may be used to refer to any degree of learned ability to speak and understand a second language, without specifying any threshold level of competence above which people are said to be "bilingual".

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 309-339. *Pacific Linguistics*, C-93, 1986.

I am assuming the basic model of intelligibility presented in Simons 1979:

total intelligibility = similarity-based intelligibility + contact-based intelligibility

In other words, intelligibility can be broken down into two components: "inherent intelligibility", due to linguistic similarity, and learned intelligibility, due to sociological factors. For modelling purposes, these factors are assumed to be independent and additive.[1]

2. INTELLIGIBILITY IN SABAH

From 1978 to 1981, the Summer Institute of Linguistics carried out a language survey of the state of Sabah, East Malaysia (formerly British North Borneo). The survey is described in detail in King and King 1984.

In the introduction to Part 2 of that volume, Carolyn Miller writes:

> The purpose of the survey was 1) to determine dialect boundaries within defined geographical boundaries comprising the entire state of Sabah, 2) to determine more precisely via lexicostatistics and intelligibility testing the degrees of intelligibility across major and minor dialect boundaries, and 3) to attempt to determine the level of understanding and the extent of the use of the national language in villages across the state.
>
> (Miller 1984)

The first phase of the survey concentrated on collecting wordlists and recording texts; the second phase involved Casad-style intelligibility testing (Casad 1974). Statistical analysis of the results demonstrates a significant correlation between cognate percentages and intelligibility testing (IT) scores, and a smaller but still significant correlation between IT scores and geographical distance.

2.1 Survey design and the nature of the data

A *correlation coefficient* (r) is an indicator of "the degree of association of strength of relationship between two variables" (Kirk 1978). When the relationship between a pair of variables is perfect (i.e. one value is perfectly predictable in terms of the other), r = 1. When the two variables are totally unrelated, r = Ø.

The Pearson product-moment correlation coefficient is the most commonly used index of correlation. It measures the strength of the relationship between two quantitative variables, e.g. IT scores and cognate percentages.

For the full set of data from the Sabah survey (790 cases), IT scores and cognate percentages are related with a Pearson correlation coefficient r = 0.663. Since r is positive, we know that the higher the cognate percentage between two dialect groups, the greater their ability to understand each other is likely to be. The statistical measurement agrees with our intuitive expectation. But in order to interpret the statistics in any meaningful way, we must know quite a bit about the units of data, how they were collected, and what these measurements represent.

Cognate percentages were calculated by computer based on a 327-word subset of the S.I.L. Philippines wordlist. Wordlists were collected from some 325 villages across the state, plus a few more from Sarawak and the Philippines.

Intelligibility testing, using the technique described in Casad 1974, was carried out at 143 of these 325 villages, and at perhaps a dozen other villages where no wordlist had been collected. A group of subjects (ideally 10) was chosen in each village, and the individual scores for each test were averaged to determine the group score (expressed as a percentage).

Constraints of time, energy and attention span forced the survey team to limit the number of tapes tested at any one village to seven. One of these was the hometown tape, one was a national language tape, and five were from other vernacular dialects. The national language score is not relevant to this study, and is not included in the data base; this leaves a total of 790 cases, i.e. 790 pairs of corresponding IT scores and cognate percentages.

Intelligibility testing was used primarily to check the language boundaries presented in Smith 1984, which were based purely on lexicostatistic relationships. This goal determined how test points were chosen and which tapes were tested at each point.

Generally, no testing was done between villages whose wordlists were more than 90% cognate, unless sociological factors made reduced intelligibility plausible (e.g. the Muslim Ida'an with the non-Muslim Begahak). Such cases were relatively rare. Very little testing was done when cognate percentages were below 70%, unless geographic proximity indicated that language learning was a strong possibility; and almost no testing was done in cases where cognate percentages fell below 50%.

In most cases, the test tapes played in each village were recorded within 50km. of that village. Testing at greater distances was done only between related dialects or closely related languages, e.g. 70% cognate or closer. On the other hand, languages much more distantly related were tested if the geographical distance between them was small. These facts tend to weaken the expected negative correlation between distance and intelligibility.

2.2 Intelligibility and lexical similarity

As mentioned above, the correlation coefficient between intelligibility (INT) and lexical similarity (LEX) over the full data set of 790 cases is $r = 0.663$. The square of this figure, $r^2 = 0.4398$, has a more intuitive interpretation. From the formula for r, we can show that r^2 is equivalent to the percentage of variation in one variable that is explainable by the variation in the related variable. In other words, 44% of the variation in IT scores can be explained by the corresponding variation in cognate percentages.

It is helpful to compare the results for the Sabah data with those from other similar studies. Simons (1979) analysed the correlation between INT and LEX for 10 different surveys in various parts of the world. These 10 represented all the studies Simons could find, prior to 1977, where both INT and LEX had been measured. The results of his analysis are shown in Figure 1-a. The corresponding values for the full data set from Sabah are shown in Figure 1-b. The meaning of the regression equations will be discussed below.

Study	N	Corr	%EV	Regression Equation
Biliau	9	.425	18.1	INT = 0.28 LEX + 66.3
Buang	21	.702	49.3	INT = 0.81 LEX - 12.4
Ethiopia	30	.846	71.6	INT = 1.22 LEX - 30.5
Iroquois	14	.813	66.0	INT = 1.52 LEX - 76.9
Mazatec	19	.807	65.1	INT = 1.77 LEX - 81.5
Polynesia	77	.864	74.6	INT = 1.59 LEX - 67.2
Siouan	25	.805	64.9	INT = 4.39 LEX - 336.0
Trique	15	.765	58.5	INT = 1.41 LEX - 41.3
Uganda	10	.905	81.8	INT = 1.33 LEX - 52.2
Yuman	25	.983	96.6	INT = 2.04 LEX - 106.2
Average	24	.791	64.6	INT = 1.05 LEX - 15.4

Figure 1-a: Ten studies from Simons 1979; INT vs. LEX

Study	N	Corr	%EV	Regression Equation
Sabah	790	.663	43.98	INT = 1.03 LEX - 3.28

Figure 1-b: Full raw data from Sabah survey, INT vs. LEX

Key:

N = number of cases
Corr = correlation coefficient, r
%EV = percentage of explained variation, r^2
Regression Equation = formula for predicting INT from LEX; defines regression line

The Sabah data set is huge in relation to any other published study of this type: 790 cases, as compared with 245 *total* cases for the 10 studies to which Simons had access. In terms of the strength of correlation (shown by r and r^2), the Sabah data is somewhat below the average of the 10 studies.

For the purposes of this study, it was decided to eliminate the hometown IT scores (i.e. subjects' scores on the test tape from their own village) from the data set. Hometown scores were included in Simons' calculations, but they are not really the same kind of measurement as other IT scores. Our model assumes that everyone understands his own dialect perfectly (i.e. 100%). Hometown tests are not tests of intelligibility but of the test itself and the subject's ability to take it.

The hometown scores in the Sabah survey were generally quite high, ranging from 80 to 100 with a mean value of 97.1. This reassures us that, on the average, the technical quality of the tests (e.g. tape quality, stories used, question construction) and the abilities of the subjects were not a major source of testing error.

The 133 hometown scores comprise 16.8%, just over one-sixth, of our data set. The LEX value for a hometown test is always 100%, and the INT values are

generally very close to 100%. Therefore, when the data is displayed as a scattergram (as in Figure 5 below), the hometown scores form a large cluster of cases around the point (100,100). The effect of removing this cluster naturally reduces the calculated strength of correlation between LEX and INT, as shown in Figure 2.

Study	N	Corr	%EV	Regression Equation
Full data	790	.663	43.98	INT = 1.03 LEX - 3.28
Exclude HT scores	657	.568	32.28	INT = 1.17 LEX - 13.12

Figure 2: INT vs. LEX, Sabah data

How can we evaluate the strength of the relationship indicated by r = 0.568? By way of analogy, we could view the measurement of cognate percentages as a kind of aptitude test. The degree of linguistic similarity between two dialects represents the innate ability of members of one dialect group to understand speakers of the other. Lexical similarity is an imperfect but useful, and easily measured, index of linguistic similarity. Taking an intelligibility test represents a complex task to which linguistic similarity is obviously relevant. Aptitude (LEX) is one of a number of variables which determine the level of actual performance of that task (INT).

Kirk (1978:108) states that the best scholastic aptitude tests rarely achieve a correlation coefficient higher than r = 0.60 between aptitude test scores and actual academic performance. If our analogy could be extended in detail, the correlation coefficient (excluding hometown scores) r = 0.568 is very respectable for an aptitude test.

However, the actual strength of the relationship between LEX and INT in Sabah is almost certainly higher than the value of r would indicate. The correlation coefficient has been reduced by the nature of the data sample, specifically its range and distribution. These problems relate to the basic design of the survey, and may be inherent in any dialect intelligibility survey situation (see section 4 below).

Kirk (1978) states that "the restriction or *truncation* of the range of [either] variable results in a misleadingly low correlation coefficient". He points out that college aptitude scores do not correlate very highly with grade point averages in college, because the college admissions process truncates the range of data. People whose aptitude scores are low do not get in.

The design of the Sabah survey had a similar effect on the range of LEX values. Because the primary aim of the survey was to establish or verify language and dialect boundaries, very little testing was done between groups that were clearly distinct linguistically. Smith (1984) used 80% cognate as an approximate threshold value below which two speech varieties could be considered distinct languages. Thus the intelligibility testing focused on the cognate range of 60-90%; only three cases below 50% cognate were tested (see Figure 3). The data set was effectively truncated at LEX = 50%.

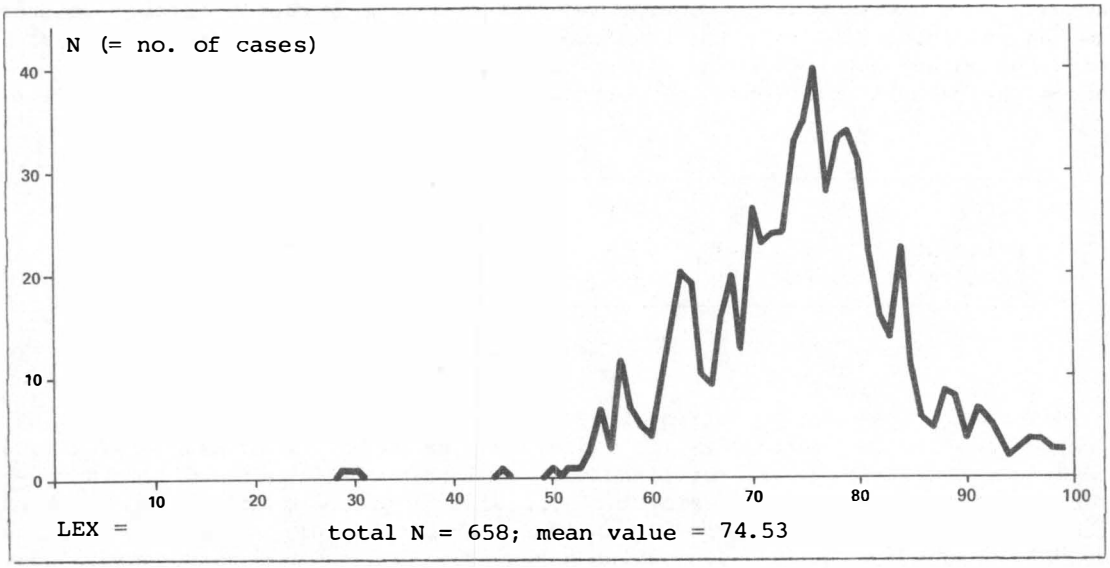


Figure 3: Distribution of LEX values for Sabah intelligibility survey

The second factor contributing to lower values of r was the *skewing* of the data, or uneven distribution of cases, particularly of the INT values. Figure 4 shows the distribution of INT. The average value of INT excluding hometown scores was 73.6%. There are slightly more cases above the mean value than below it, and there are far more occurrences of each value above 73% than of the values below that figure. The distribution of LEX is also skewed somewhat to the right (higher values), but far less so than INT.[2]

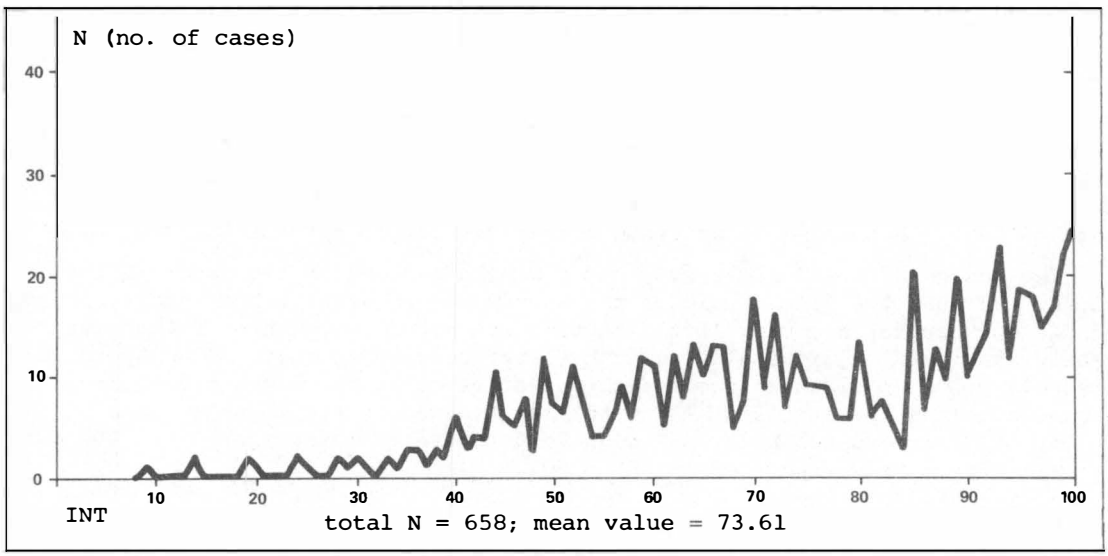


Figure 4: Distribution of INT values

The distribution of IT scores is related to the testing methodology. Following Casad 1974, simple personal experience stories were used, so average IT scores between dialects of the same language were rarely below 80%. On simple first-person narratives of this type, the ability to answer 10 questions out of 10 correctly does not necessarily indicate comprehension equaling that of a native speaker. But since no score higher than 100% is possible, the mean IT scores between related dialects appear as a dense cluster between 80% and 100%.

Another factor at work is the effect of language learning, which tends to inflate IT scores. There is no effect of anything like the same magnitude working to lower scores, so the net result is a higher frequency of high scores.

"If the distributions of [the variables] are markedly skewed, the value of r will be less than if the variables are approximately normally distributed" (Kirk 1978:113). Once again, it seems safe to predict that the actual correlation between LEX and INT in Sabah is greater than the measured value, r = 0.568.

The relationship of LEX and INT can be approximated by a linear equation of the form: INT = aLEX + b. Linear regression analysis is a technique for calculating the parameters (a and b) of this equation. The *regression line* defined by the linear equation is the line of best prediction for INT in terms of LEX. For the data set as a whole, the total difference between actual measured values in INT and the predicted values based on corresponding values of LEX (that is, the total prediction error) is minimised. Figure 5 shows a scattergram of the Sabah data (excluding hometown scores) with the associated regression line (line a), defined by the equation INT = 1.17 LEX - 13.12.

For normally distributed data, the regression line is the line of best fit (the line which passes closest to all the points in the scattergram), and r is a measure of how closely the points cluster around the line. However, we have already noted that the Sabah data is not normally distributed, and visual inspection of the scattergram shows that the regression line does not fit the points very well. This is confirmed by residual analysis, i.e. plotting prediction error against observed values of INT (see Figure 6). The rising trend in Figure 6-a indicates that the prediction error is roughly linearly dependent on the observed value of INT, and thus that the formula for the regression line does not adequately describe the data.

In our regression analysis, we have assumed that LEX was the independent variable and INT the dependent. In other words, we choose (for theoretical reasons) to predict INT from LEX rather than vice versa.

The shape of the regression line is dependent on the choice of dependent variable, and, as Figure 5 shows, the regression line for predicting LEX from INT (line b) is quite different from the first line. By visual inspection, it appears to fit the data much better than the first regression line. Figure 6-b shows the residual analysis for line b, Figure 5. The even, horizontal pattern in Figure 6-b indicates that the prediction error is random with respect to the observed value of INT, and so line b does in fact fit the data better than line a.

Line b is also a more plausible model of the actual relationship between LEX and INT than line a. Line b (INT = 3.57 LEX - 193.2) predicts zero inherent intelligibility between any two languages less than 54% cognate with each other, and full intelligibility between pairs of languages above 82% cognate. Intuitively, we would expect a higher threshold for full intelligibility, e.g. 90-92% cognate,[3] but the basic shape of line b is at least suggestive of the type of model we expect.

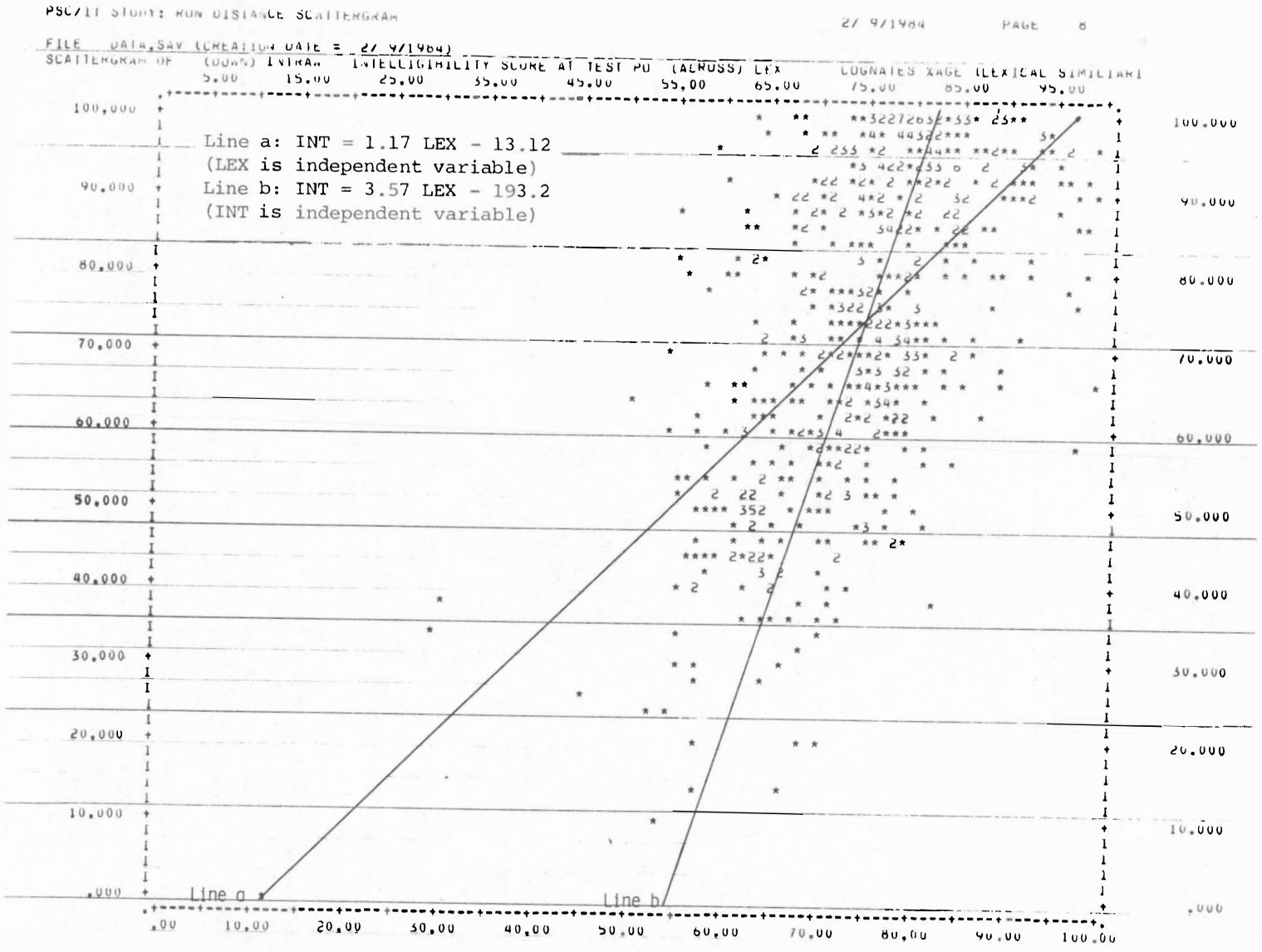


Figure 5: Scattergram of INT vs. LEX with regression lines

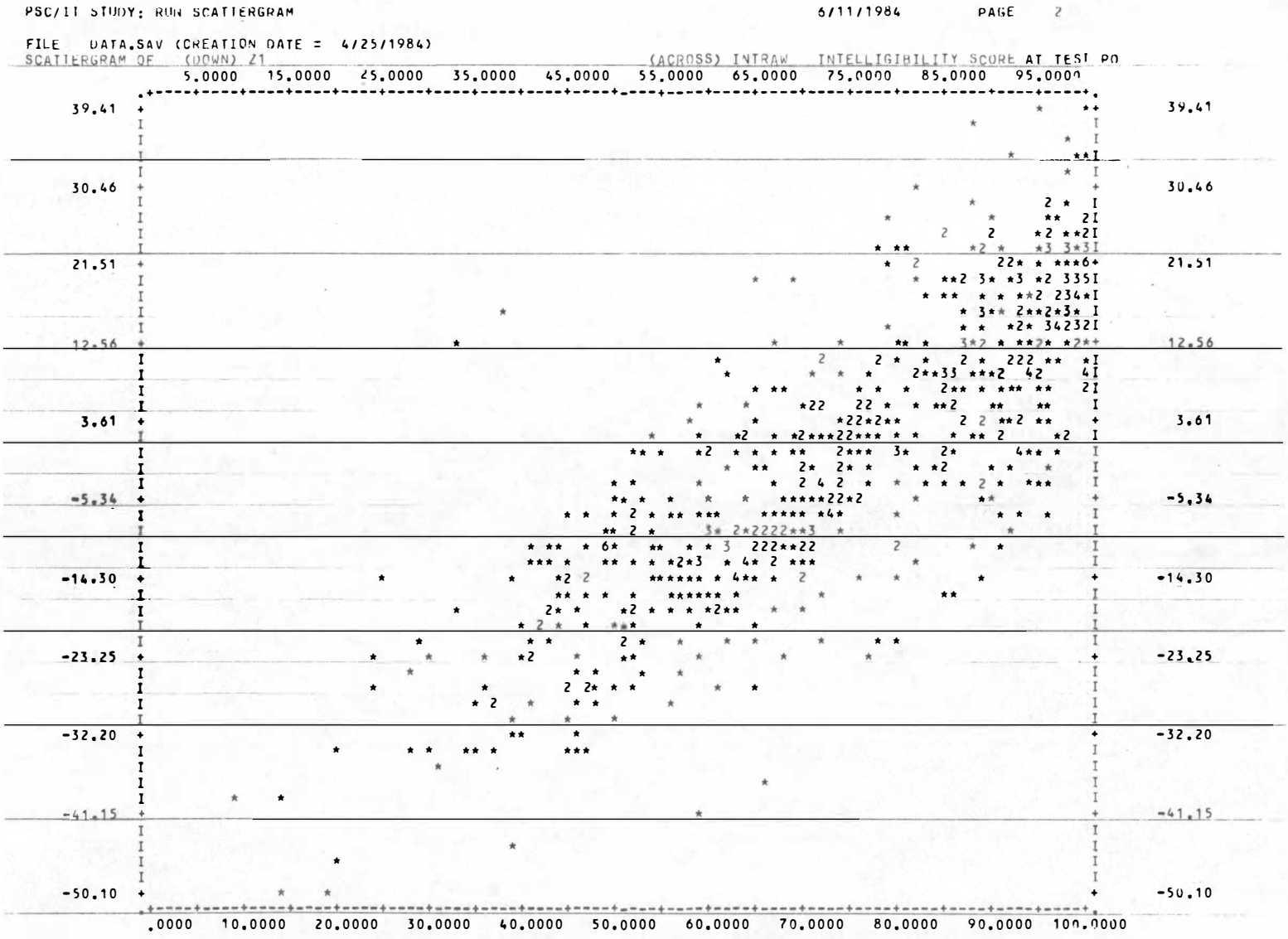


Figure 6-a: Residual analysis with LEX as independent variable. Prediction error vs. measured INT

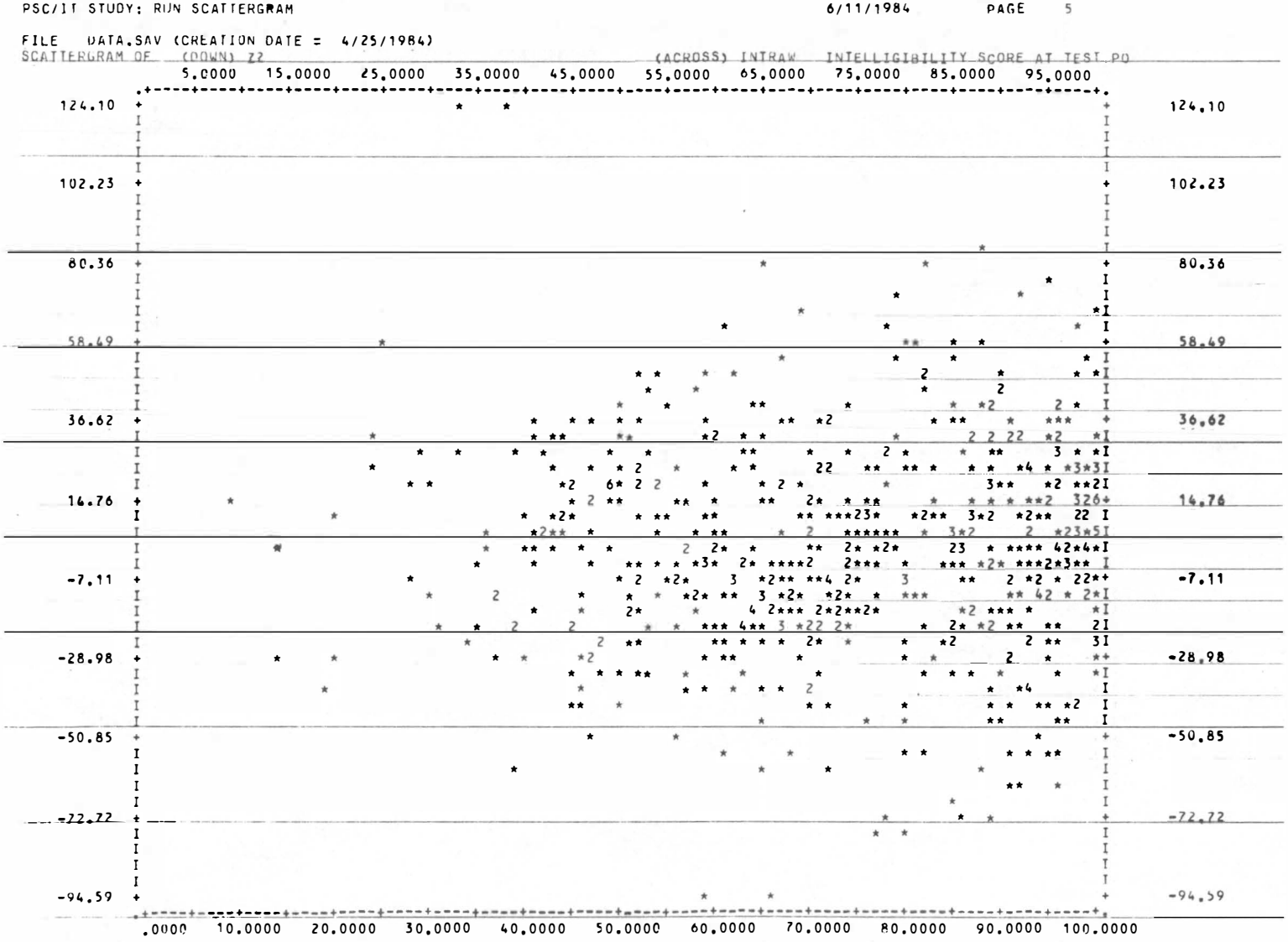


Figure 6-b: Residual analysis with INT as independent variable. Prediction error vs. measured INT

Line a, on the other hand, is much too "flat". It predicts zero intelligibility only below 11% cognate, when experience tells us that this threshold must be closer to 50%.[4]

But no theory of language would treat lexical similarity as dependent on intelligibility, so it is nonsense to say that the second line is the "correct" one. The difference in the two lines may be related to the fact that LEX is more nearly normal in distribution than INT. Both lines tell us something about our particular data set, but probably neither line tells us much about the actual relationship between LEX and INT in Sabah.

2.3 Intelligibility and distance

People who live near each other are more likely to interact than people who live far apart. We expect to find a negative correlation between distance and language contact, and therefore a negative correlation between distance and intelligibility.

2.3.1 Previous studies

Simons (1979:ch.6) shows that relative distance (relative to position in the dialect area) is a better predictor of intelligibility than absolute distance, for the dialects on Santa Cruz Island. Unfortunately, the results of that study are not comparable to ours, because he measured intelligibility on a discrete point scale (3 = full, 2 = partial, 1 = sporadic, Ø = none) rather than as a percentage.

Simons measured distance in travel time, which is clearly more relevant than raw physical distance. However, Walter and Echerd (n.d.) present a very interesting study using raw, straight-line distances measured on a map. For the Cakchiquel dialect system, intelligibility correlates very strongly with the natural logarithm of distance (see Figure 7).

Study	N	Corr	%EV	Equation
Cakchiquel	19	.96	92.16	INT = 332.77 - 93.17 ln(DIST)

Figure 7: Cakchiquel data from Walter and Echerd (n.d.) LEX vs. INT

The correlation coefficient r is an astonishing 0.96, equivalent to 92% explained variation. The authors apologise for the crudeness of the measurement (straight-line measurements on a map), but the results for the Cakchiquel data leave no room for improvement! Indeed, their model (using only distance) seems to be more accurate than intelligibility testing itself.

An interesting feature of that model is that it predicts full intelligibility for any pair of villages less than 12.6km. apart. Walter and Echerd suggest that this distance is the radius of an "interaction zone", defined as "that geo-linguistic zone in which a person moves with sufficient freedom and regularity so that he characteristically attains and maintains complete intelligibility of communication with all those (with) whom he comes in contact."

2.3.2 The Sabah data

In the present study, simple straight-line measurements on a map are used as the distance measure. The data set is reduced to include just the three major indigenous language families of Sabah: Dusunic, Murutic and Paitanic (see language map in Figure 8). However, because the bulk of the survey was focused on these groups, this subset includes 700 cases, or 88.6% of the full data set.

Taking the data as a whole, we find the expected negative correlation between distance and intelligibility but the correspondence is weak (see Figure 9).

Study	N	Corr	%EV	Equation
Full data	700	-.444	19.71	INT = 86.08 - 1.02 DIST
Excl. Hometown scores	592	-.316	9.97	INT = 81.03 - 0.73 DIST

Figure 9: Correlation between distance and intelligibility in Sabah
(Dusunic, Murutic and Paitanic families only)

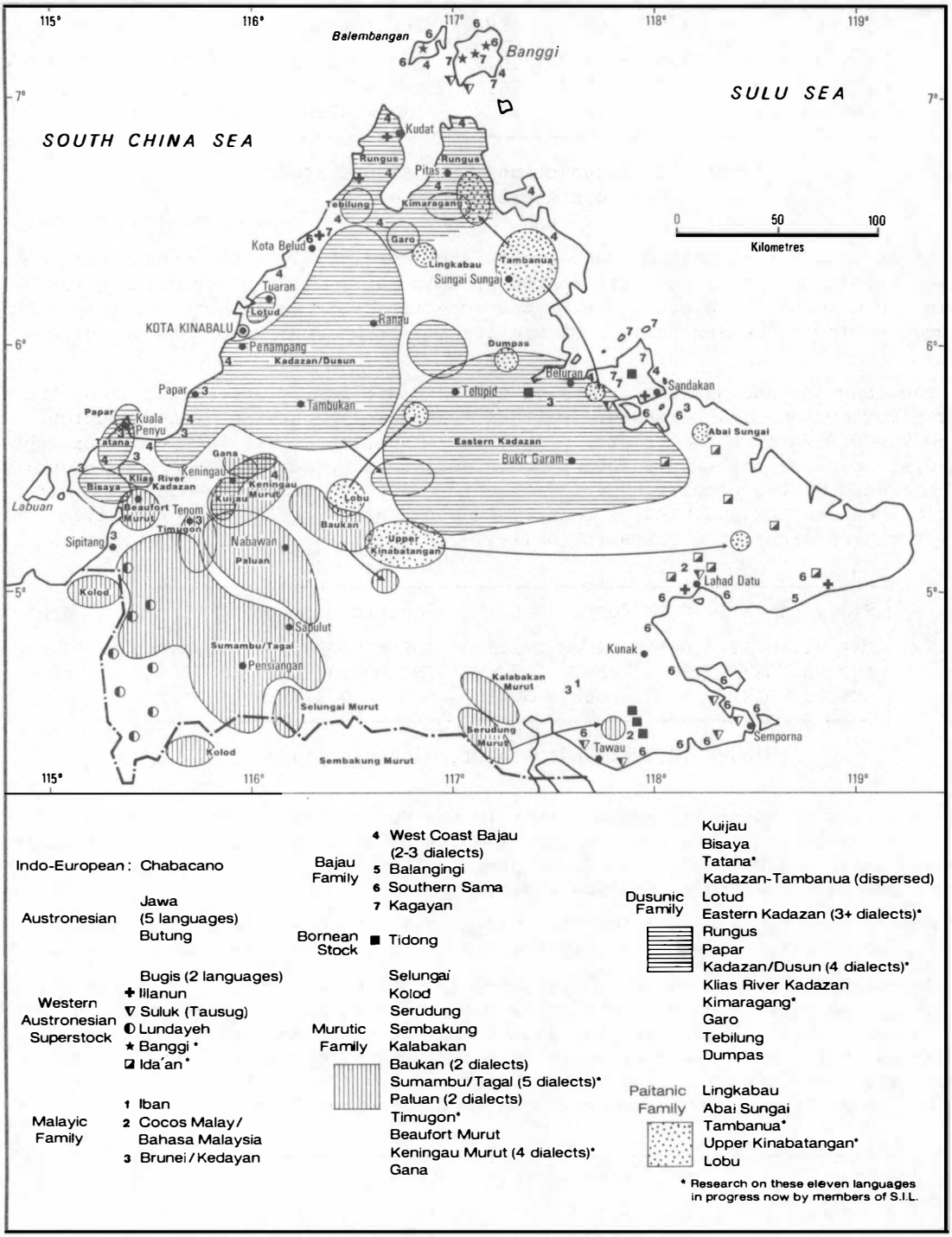
The weak correlation between INT and DIST is partly a result of the survey design, as discussed above, and partly due to the mixture of groups from different language families in many areas of the state.

When we take various subsets of the data, more interesting patterns emerge. For the Murutic subset (both Speaker and Hearer belonging to Murutic language groups), distance is a much better predictor of intelligibility than lexical similarity is. Figure 10 shows that 44% of the variation in INT can be explained as a function of the variation in DIST, compared with only 15% for LEX.

Study	N	Corr	%EV	Equation
INT vs. DIST	96	-.664	44.09	INT = 92.43 - 1.68 DIST
INT vs. LEX	96	.396	15.68	INT = 0.96 LEX + 5.32
LEX vs. DIST	96	-.349	12.20	LEX = 102.43 - 2.97 DIST

Figure 10: Murutic subset, distance study
(Hometown scores excluded)

For the Dusunic language family (Figure 11), the correlation between INT and DIST is stronger than that shown in Figure 9, line 2, for the data set as a whole, but still not very high. LEX is a slightly better predictor of intelligibility than is DIST, but neither LEX nor DIST alone can account for even 30% of the variation in INT. This effect is probably due to the fact that sociological factors, e.g. relative prestige differences, are more extreme in the Dusunic family than in the other language families of Sabah.



Indo-European: Chabacano

Austronesian
- Jawa (5 languages)
- Butung

Western Austronesian Superstock
- Bugis (2 languages)
- ✚ Illanun
- ▽ Suluk (Tausug)
- ◐ Lundayeh
- ★ Banggi *
- ◪ Ida'an *

Malayic Family
- 1 Iban
- 2 Cocos Malay/ Bahasa Malaysia
- 3 Brunei/Kedayan

Bajau Family
- 4 West Coast Bajau (2-3 dialects)
- 5 Balangingi
- 6 Southern Sama
- 7 Kagayan

Bornean Stock
- ■ Tidong

Murutic Family
- Selungai
- Kolod
- Serudung
- Sembakung
- Kalabakan
- Baukan (2 dialects)
- Sumambu/Tagal (5 dialects)*
- Paluan (2 dialects)
- Timugon*
- Beaufort Murut
- Keningau Murut (4 dialects)*
- Gana

Dusunic Family
- Kuijau
- Bisaya
- Tatana*
- Kadazan-Tambanua (dispersed)
- Lotud
- Eastern Kadazan (3+ dialects)*
- Rungus
- Papar
- Kadazan/Dusun (4 dialects)*
- Klias River Kadazan
- Kimaragang*
- Garo
- Tebilung
- Dumpas

Paitanic Family
- Lingkabau
- Abai Sungai
- Tambanua*
- Upper Kinabatangan*
- Lobu

* Research on these eleven languages in progress now by members of S.I.L.

Figure 8: Languages of Sabah

Study	N	Corr	%EV	Equation
INT vs. DIST	325	-.447	20.02	INT = 89.52 - 1.16 DIST
INT vs. LEX	325	.514	26.37	INT = 1.40 LEX - 30.98
LEX vs. DIST	325	-.505	25.49	LEX = 97.69 - 1.89 DIST

Figure 11: Dusunic subset, distance study
(Hometown scores excluded)

It is interesting that distance correlates more highly with cognate percentage than with intelligibility in the Dusunic family. This correlation reflects the extensive dialect chaining characteristic of Dusunic groups, with each village tending to be lexically most similar to its nearest geographical neighbours.

The language and dialect groups of the Paitanic family are spread along the major river systems in the eastern part of the state. Patterns of interaction follow the course of these rivers. For this reason, as Figure 12 shows, straight-line distances are not very relevant to intelligibility scores among this group; LEX is a much better predictor of INT than is DIST. On the other hand, the correlation between LEX and DIST is relatively high, again reflecting extensive dialect chains within the Paitanic family.

Study	N	Corr	%EV	Equation
INT vs. DIST	86	-.464	21.52	INT = 82.98 - 0.60 DIST
INT vs. LEX	86	.582	33.84	INT = 0.94 LEX - 0.74
LEX vs. DIST	86	-.608	36.95	LEX = 99.37 - 1.31 DIST

Figure 12: Paitanic subset, distance study

Rungus is a Dusunic language spoken in the Kudat Division in the northern part of the state. The Rungus are the majority population group in most of their language area, and Rungus is the local prestige language, the church language even in several non-Rungus speaking areas.

When Rungus tapes are tested with non-Rungus subjects, distance is a much better predictor for intelligibility than is lexical similarity (see Figure 13-a).

However, for Rungus subjects listening to non-Rungus tapes, LEX is a better predictor than DIST (Figure 13-b). This suggests that Rungus speakers do not tend to learn other languages; they understand dialects that are linguistically similar to their own. However, other language groups in the area tend to learn Rungus, and those groups living closest to the Rungus learn it best. This is exactly the pattern we would expect for a local prestige language.

Study	N	Corr	%EV	Equation
INT vs. DIST	20	-.707	49.99	INT = 69.61 - 1.22 DIST
INT vs. LEX	20	.498	24.78	INT = 1.27 LEX - 34.98

Figure 13-a: Rungus speaker, non-Rungus hearer

Study	N	Corr	%EV	Equation
INT vs. DIST	31	-.318	10.09	INT = 66.68 - 0.59 DIST
INT vs. LEX	31	.614	37.75	INT = 1.08 LEX - 18.56

Figure 13-b: Non-Rungus speaker, Rungus hearer

Penampang Kadazan (the subdialect of Coastal Kadazan spoken in western Penampang District) is also a high-prestige dialect. Like the Rungus, the Penampang Kadazan dialect group has a high awareness of its identity as a group. It is larger in population than the Rungus, and is the most politically influential of the indigenous language groups in Sabah. Penampang Kadazan is used by the Catholic church in many areas of the state, and is also used in newspapers, magazines, radio, etc.

The pattern of intelligibility for other groups listening to Penampang Kadazan tapes (Figure 14) is even more striking than in the Rungus case. In terms of percentage of explained variation, distance is twice as good a predictor of intelligibility as lexical similarity is. However, no pattern emerges from what little data are available (only 13 cases) for Penampang Kadazan listening to other dialects.

Study	N	Corr	%EV	Equation
INT vs. DIST	44	-.802	64.39	INT = 94.79 - 1.90 DIST
INT vs. LEX	44	.579	33.58	INT = 1.83 LEX - 64.28

Figure 14: Penampang Kadazan speaker, non-Penampang Kadazan hearer

Another interesting case study is Kuijau, a Dusunic language heavily mixed with Murutic vocabulary. The Kuijau are a lower-prestige and somewhat scattered group living among the various Murutic groups of the Keningau District. The IT scores for Kuijau subjects listening to other languages show a surprising *positive* correlation between DIST and INT; in other words, the farther away a group lives, the better the Kuijau understand them (Figure 15).

Study	N	Corr	%EV	Equation
INT vs. DIST	16	.444	19.74	INT = 0.86 DIST + 66.46
INT vs. LEX	16	-.106	1.13	(no significant relationship)

Figure 15: Kuijau hearer, non-Kuijau speaker

This pattern is partly due to dialect geography, partly the result of the survey design. The closest neighbours of the Kuijau are speakers of various Murutic dialects; but the Kuijau are linguistically more similar to the Central Dusun of neighbouring Tambunan District. The Kuijau have contact with Dusun immigrants from Tambunan, and often refer to themselves as Dusun.

Kuijau subjects were tested with geographically close but linguistically distant Murutic tapes; and with geographically distant but linguistically closer -

and highly prestigious - Central Dusun and Penampang Kadazan tapes. Thus the expected relationship between IT scores and distance was reversed.

2.4 Summary

Linguistic similarity is obviously an important factor in predicting how well members of one dialect group will understand speakers of another group. Lexical similarity (cognate percentage) is a useful index of linguistic similarity, and the Sabah data exhibit the expected correlation between cognate percentage and intelligibility. However, simple regression analysis of the data cannot determine the precise mathematical relationship between these two measurements, for reasons that will be discussed further in section 4 below.

Geographical distance is related to language contact, and is found to be a significant factor in situations where bilingualism is a major component of the measured intelligibility. Other social factors affecting intelligibility in particular situations in Sabah will be discussed in the following section.

3. PATTERNS OF BILINGUALISM: TWO CASE STUDIES

As shown in section 2.3 above, there are some subsets of the data for which distance is a better predictor of intelligibility than is lexical similarity. This is particularly true in the case of prestige dialects, such as Rungus and Penampang Kadazan.

The distance between two groups is related to the opportunity for contact between them, thus the component of intelligibility due to social factors rather than the component due to linguistic similarity is in focus here. The relatively high correlation between distance and measured intelligibility in some sets of data suggests that the IT scores reflect not only inherent intelligibility but also a significant amount of learned comprehension or bilingualism.

In studying inherent intelligibility, it is appropriate to use average IT scores, because inherent intelligibility is assumed to be fairly uniform throughout a speech community. However, average IT scores are useless for investigating bilingualism. There is often a wide range of scores among different segments of the population, and that is precisely the phenomenon that we want to investigate. What members of group X understand language Y and to what extent?

To get this information, we must use individual IT scores. This greatly increases the volume of data, and the amount of work involved in processing and analysing the data. Two case studies are discussed here: outsiders' comprehension of oral texts in Rungus and Penampang Kadazan. Even for these two examples, it has not been possible to process all the data available. Out of roughly 200 non-Rungus individuals tested with Rungus tapes, a representative sample of 88 individuals was chosen for this part of the study. Out of roughly 440 non-Penampang Kadazan subjects tested with Penampang Kadazan texts, a sample of 192 individuals was selected. For each sample group, an attempt was made to include at least one village from each dialect group where Rungus or Penampang Kadazan was tested.

3.1 Correlation study

Each subject's score on the Casad-style intelligibility test is taken here as a measurement of his ability to understand Rungus or Penampang Kadazan, as the case may be. Therefore, the INT values for each subject range from 0 to 10, representing the number of questions answered correctly.

Various other measurements were included in this phase of the study, including: LEX, the cognate percentage between the subject's dialect and the test dialect; DIST, the straight-line distance from the subject's village to the village where the test tape was recorded; MALAY, the subject's individual score on the Bahasa Malaysia test, ranging from 0 to 10; AGE, the subject's age in years; SEX, subject's gender; EDUC, amount of schooling in years; TRAVEL, extent of travel outside home language area, on a scale of 1 to 5; BIRTH, relative distance of birthplace from present residence, i.e. from village where the subject was tested, on a scale of 1 to 3.

For this part of the study, the Spearman rank-order coefficient r_s is used as a measure of correlation. Unlike the Pearson product-moment correlation coefficient used above, the rank-order coefficient is a non-parametric measurement which does not assume that the data is normally distributed, and which can be applied to simple ranking scales as well as pure quantitative measurements. As with Pearson's r, the values of r_s range from -1 to 1, with $r_s = Ø$ indicating that no relationship is measurable between the two variables.

Correlation analysis is useful for picking out linear relationships among the variables being studied. Figure 16 shows the correlation coefficients (r_s) and associated measures of significance for pairs of variables which seem most strongly related in the Penampang and Rungus data.

Significance[5] is a measurement of the likelihood that a particular pattern is purely accidental. For example, a significance value of .01 indicates that there is one chance in a hundred that an observed association is purely random; or, 99% certainty that it reflects some real characteristic of the population from which the data was drawn. A significance value of zero indicates perfect confidence, i.e. zero probability that the pattern is due merely to chance. A value of 1 is the worst possible case; it means that there is no room for doubt that the observed pattern is accidental.

In general, the significance value is closer to zero, i.e. better, for stronger correlations (larger values of r or r_s) and for larger data sets. A correlation of $r_s = .25$ may be significant in the Penampang study, with 191 cases, but not in the Rungus study with only 87 cases.

Simons (1977) recommends using a .10 level of significance (i.e. a 90% confidence level) for determining significant differences in lexicostatistics. For the purposes of this study, I would consider any correlation with a significance value below .01 as definitely significant, and any below .10 as being worthy of further investigation.

	Penampang Kadazan		Rungus	
VARIABLES	CORR	SIG	CORR	SIG
INT/LEX	---	---	.3853	.001
INT/DIST	-.4788	.001	-.4719	.001
INT/MALAY	.2786	.001	---	---
INT/EDUC	.2907	.001	---	---
INT/TRAVEL	.2286	.002	(-.1986	.067)
LEX/DIST	---	---	-.3483	.001
MALAY/EDUC	.5178	.001	.7433	.001
AGE/EDUC	-.6097	.001	-.5499	.001
MALAY/AGE	-.3379	.001	-.3959	.001
DIST/EDUC	(-.1392	.055)	(.2053	.056)

Figure 16: Correlation and significance values for Penampang Kadazan and Rungus bilingualism studies

Key:

CORR = Spearman rank-order correlation coefficient
SIG = significance (two-tailed test)
INT = individual IT score
LEX = cognate percentage between subject's dialect and test dialect
DIST = linear distance between test point and village where test tape was recorded
MALAY = individual score on Bahasa Malaysia test
EDUC = years of formal education
TRAVEL = extent of travel outside subject's dialect area
AGE = subject's age in years
( ) = marginally significant relationship

The first two lines of Figure 16 confirm the results of section 2.3, which were based on average INT scores. The correlation between INT and DIST in the Penampang data is not so striking here, and there is now no significant correlation between INT and LEX. This is because using individual IT scores allows differences in age, education, travel, etc. to overshadow the relatively smaller effect of lexical similarity. However, for the Rungus test, the INT vs. LEX and INT vs. DIST correlations were roughly the same as those computed from average IT scores.

We are primarily interested in factors which correlate highly with INT. However, the strongest patterns in the data involve three highly interrelated variables: age, amount of education, and ability in the national language, Bahasa Malaysia (lines 7, 8 and 9).

The high correlation between ED and MALAY tells us that the more schooling a person has, the better he will understand Malay. This pattern is especially striking in the Rungus study (r_s = .7433). The negative correlation between AGE and ED says that the younger people are generally better educated than their elders. This is the strongest single relationship in the Penampang testing sample, r_s = -.6097. And the negative correlation between AGE and MALAY says that, on the average, young people understand Malay better than their elders.

A crucial difference between the two studies shows up in the correlations of INT with ED and MALAY (lines 3 and 4). Penampang Kadazan is a state-wide prestige language, used on the radio, in newspapers, at political rallies, etc. A significant body of Penampang Kadazan literature exists. Among non-Kadazans, the better educated and more upwardly mobile (those who tend to speak Malay better) understand the Penampang dialect better.

The prestige of Rungus, by contrast extends over a fairly limited area. People learn Rungus at weekly markets and in other traditional contexts of social interaction, not through the mass media. Only a very small body of Rungus literature exists, and it is not widely distributed even within the Rungus community. Thus, education and ability to understand Malay are irrelevant to a person's ability to understand Rungus.

A related difference is apparent in the relationship between INT and TRAVEL (line 5). The weak but fairly significant positive correlation in the Kadazan study (r_s = .2286, significance = .002) indicates that people who have travelled further from their native language areas tend to understand Penampang Kadazan better than those who stay at home. The marginally significant *negative* correlation in the Rungus study (r_s = -.1986, significance = .067) suggests that those who stay at home tend to understand Rungus better than those who travel.

This difference could be related to the urbanisation of Sabah, one of the most important population trends in the state today. The Penampang Kadazan dialect area is contiguous to the state capital, Kota Kinabalu. Of the three major towns in Sabah, the capital naturally exerts the strongest attraction on people from outlying districts. For most of the people in the Penampang test sample, when they travel, they travel towards Penampang.

The Rungus area, on the other hand, is one of the least developed areas in the state. The non-Rungus who leave their own areas have little incentive to go north towards the Rungus area, and as they go south towards the capital, they are cut off from contact with the Rungus language.

Finally, there is a marginally significant correlation in both studies between DIST and ED (line 10). This suggests that people living closer to Penampang (and thus to the capital) tend to get more education than those farther in the interior (r_s = -.1392). People who live closer to the Rungus area, i.e. farther north, tend to get *less* education than those who live to the south (r_s = 0.2053).

In addition to the variables listed in Figure 16, data were collected for each subject about his or her spouse's first language. A simple scale was used to rate the degree of difference between the spouse's language and the subject's mother tongue: 1 if both were native speakers of the same dialect, 2 if they spoke different dialects or languages within the same language family (e.g. both Dusunic or both Murutic), 3 if they spoke dialects from different language families (e.g. Kadazan and Bajau, or Murut and Malay).

Somewhat surprisingly, no correlation was found between linguistic diversity in the marriage and ability to understand either Rungus or Kadazan. However, in both studies it appears that people who marry outside their own language group tend to be better educated than those who marry within the group (Rungus study: r_s = .3596, sig = .002; Penampang study: r_s = .2693, sig = .001). There is also a tendency for better educated people in both sample groups to travel more widely than their less-educated counterparts, and a weak tendency for people who marry outside the language group to be more widely travelled than those who marry within the group. Finally, in both studies, subjects who married outside their own language group scored higher on the Malay test than those who married within the group (Rungus study: r_s = .3277, sig = .005; Penampang study, r_s = .2092, sig = .018).

3.2 Tabulation of the data

Correlation analysis can reveal linear trends in the data, but a simple tabulation of the data is helpful in interpreting these trends, and in finding other, non-linear, relationships.

One obvious pattern which correlation analysis could not reveal is the fact that men, on the average, understand Penampang Kadazan better than women.[6] Figure 17 shows the breakdown of scores by sex for both studies; notice that in the Rungus test, there was virtually no difference in scores between the sexes.

	Penampang Kadazan			Rungus		
SEX	N	MEAN	STD.DEV.	N	MEAN	STD.DEV.
Male	111	7.14	1.900	57	5.46	1.864
Female	81	6.54	2.060	31	5.37	2.152
Total	192	6.89	1.965	88	5.43	1.970

Figure 17: Breakdown of intelligibility scores by sex for the Penampang Kadazan and Rungus studies

Key:

N = number of cases
MEAN = average score
STD.DEV. = standard deviation

The difference between men's and women's scores in the Penampang study may be related to the correlation mentioned in section 3.1 between extent of travel and ability to understand Kadazan. In both sample groups, men on the average have travelled more widely outside their home language area than women.[7] In the Rungus study, men are also better educated than women (average 3.79 years for men, 2.68 years for women);[8] but we have already seen that there is no correlation between years of education and ability to understand Rungus. In the Penampang study, the difference in education is not statistically significant (the mean for women being slightly higher than that for men). Thus the difference between men's and women's scores on the Kadazan test are not related to educational differences.

Figure 18 shows a breakdown of scores by occupation. The "agricultural" category includes farmers and fishermen; in these two samples, most people in this category are rice farmers. "Government employee" includes village headmen, native chiefs, teachers, community development officers and civil servants (all those tested were also residents of the villages where data were collected and native speakers of the dialect spoken in their village). "Other" includes small business men, students, unemployed, etc.

	Penampang Kadazan			Rungus		
OCCUPATION	N	MEAN	STD.DEV.	N	MEAN	STD.DEV.
Agriculture	65	7.01	1.882	47	5.55	1.877
Government	23	7.59	2.175	15	4.37	2.254
Other	93	6.69	2.031	26	5.81	1.647

Figure 18: Breakdown of intelligibility scores by occupation for the Penampang Kadazan and Rungus studies

The interesting comparison here is in the scores of government employees. This group did better than either farmers or "others" on the Penampang Kadazan test, but scored lower than either of the other categories on the Rungus test.[9]

Figure 19-a presents a breakdown of INT scores based on how extensively a subject had travelled outside his own language area. Category 1 indicates that the subject had never left the language area; 2 that he/she had travelled only to neighbouring districts; 3 indicates extensive travel within the state, e.g. from west coast to east coast; 4 indicates travel outside the state, generally to Sarawak, West Malaysia or Singapore; 5 indicates that the subject had lived for extended periods of work or study outside his/her own language area (whether in Sabah or elsewhere).

These results confirm the correlation findings (see Figure 16, line 5) showing that the more widely travelled subjects understood Penampang Kadazan better, while those who had travelled less understood Rungus better. In category 5 (which deals with residence rather than travel) is ignored, the trend lines in Figure 19-b are monotonic in both studies, and strictly monotonic for the Penampang study.

	Penampang Kadazan			Rungus		
TRAVEL	N	INT	STD.DEV.	N	INT	STD.DEV.
1	23	6.17	2.278	19	5.68	1.407
2	90	6.67	1.774	46	5.68	1.872
3	50	7.14	2.124	8	4.44	2.468
4	10	7.65	1.718	3	4.17	1.312
5	17	7.65	1.845	10	4.60	2.289

Figure 19-a: Breakdown of INT by travel for the Penampang Kadazan and Rungus studies

Key:

N = number of cases
INT = average of individual IT scores
STD.DEV. = standard deviation
TRAVEL = extent of travel outside subject's home language area:
- 1 = never left language area
- 2 = travel only to neighbouring districts
- 3 = state-wide travel
- 4 = travel outside Sabah
- 5 = live for one year or more outside home language area

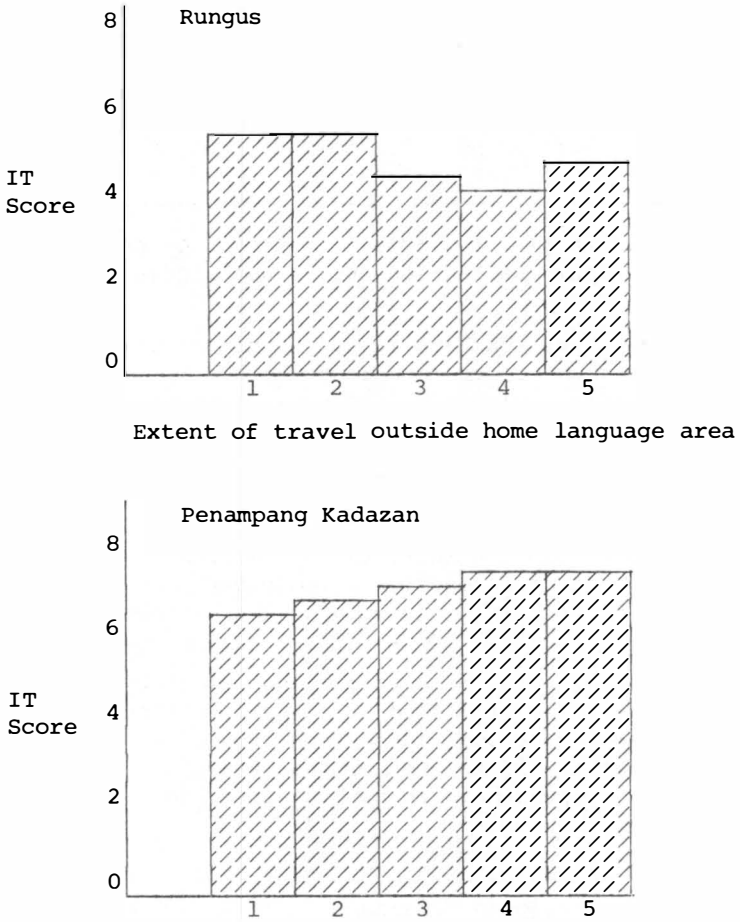


Figure 19-b: IT score vs. travel for Rungus and Penampang Kadazan studies

Correlation analysis revealed no linear relationship between AGE and INT for either test. Figure 20-a shows a breakdown of scores by age and sex, and Figure 20-b shows the same information as a graph. There is no clear trend for the sample as a whole in either graph, but in both studies women tend to score highest in middle age. In both studies, the highest mean score for women is in the 30-39 age group, and women less than 30 score higher than women over 50.

	Penampang Kadazan		Rungus	
AGE	N(m/f)	INT(men/women)	N(m/f)	INT(men/women)
10-19	33 (13/20)	6.97 (7.58/6.58)	14 (7.7)	5.61 (5.71/5.50)
20-29	60 (33/27)	6.89 (7.05/6.70)	24(14/10)	5.71 (5.46/6.05)
30-39	33 (20/13)	7.11 (7.15/7.04)	15 (12/3)	5.13 (4.54/7.50)
40-49	32 (22/10)	6.59 (6.57/6.65)	17 (13/4)	6.20 (6.38/5.63)
50-59	17 (10/7)	6.94 (8.10/5.29)	7 (4/3)	4.57 (5.25/3.67)
60-69	14 (10/4)	6.50 (6.85/5.63)	9 (5/4)	4.34 (5.50/2.88)
over 70	3 (3/0)	8.00 (8.0/-)	2 (2/0)	4.25 (4.25/-)

Figure 20-a: Breakdown of INT by age group for the Penampang Kadazan and Rungus studies

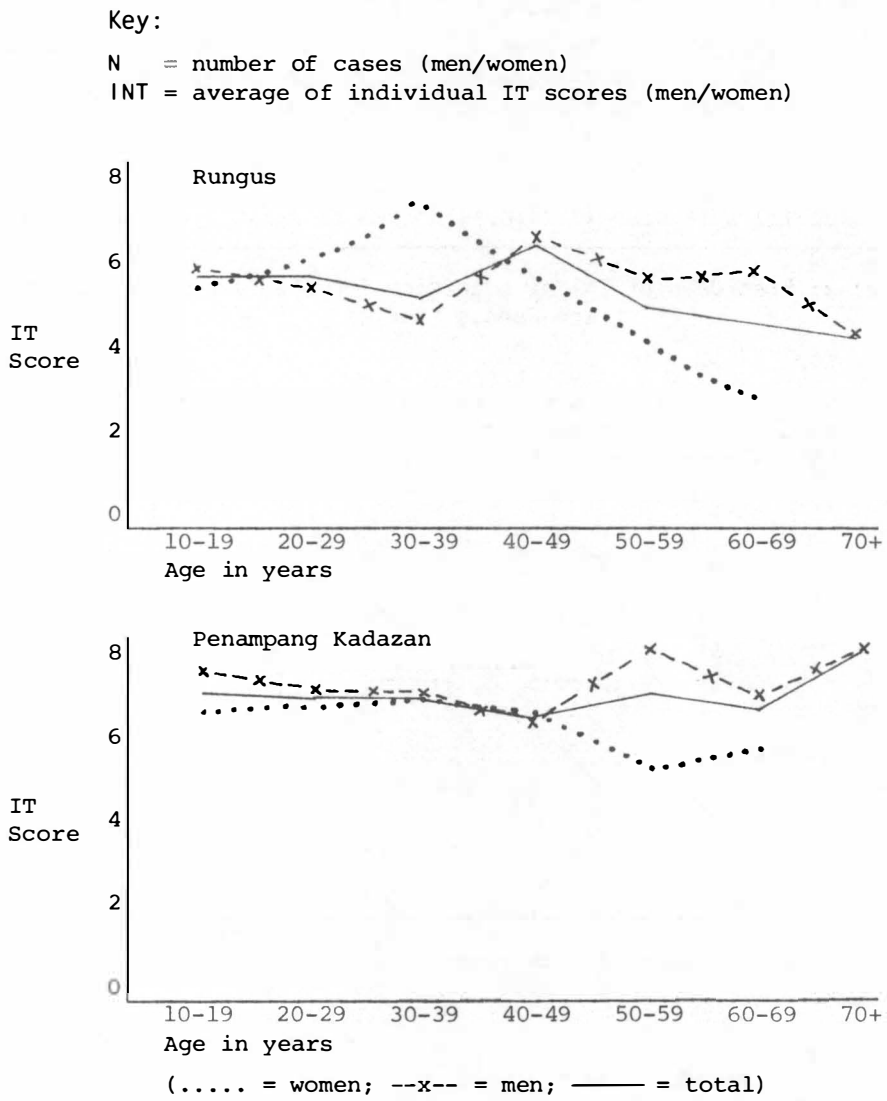


Figure 20-b: IT score vs. age group for Rungus and Penampang Kadazan studies

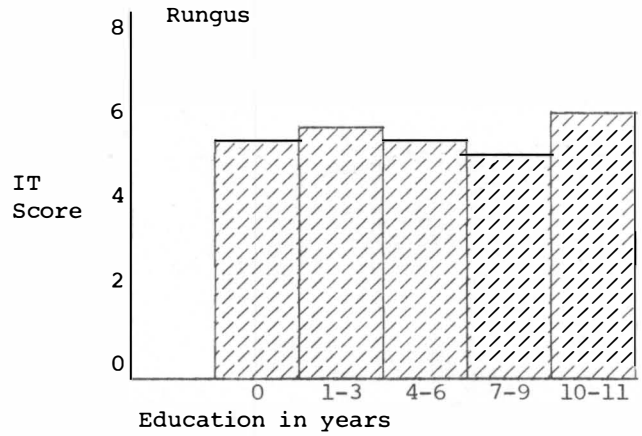
Figure 21-a shows a breakdown of scores by amount of education (in years). The graph in Figure 21-b confirms the positive correlation mentioned in Section 3.1 between INT and ED for the Penampang Kadazan test. No trend is apparent in the Rungus study. It is apparent that the subjects who took the Penampang Kadazan test were, on the average better educated than those who took the Rungus test. A higher percentage of those in the Rungus study had no education (50% vs. 34%), and a higher percentage of those that had been to school never got beyond Primary Six (66% vs. 56%). This is a significant difference between the two sample sets, and makes a rigorous comparison of specific results between the two studies more difficult. However, it should not affect the interpretation of trends within each study.

	Penampang Kadazan		Rungus	
EDUC	N(m/f)	INT(men/women)	N(m/f)	INT(men/women)
0	66 (35/31)	6.27 (6.74/5.73)	44 (26/18)	5.46 (5.75/5.03)
1-3	17 (14/3)	6.74 (6.78/6.50)	3 (2/1)	5.50 (4.75/7.00)
4-6	53 (33/20)	6.92 (7.20/6.48)	26 (18/8)	5.39 (5.33/5.50)
7-9	45 (25/20)	7.63 (7.56/7.73)	9 (5/4)	5.06 (4.10/6.25)
10-11	9 (2/7)	7.39 (8.75/7.00)	6 (6/0)	5.92 (5.92/-)
12-over	2 (2/0)	8.50 (8.50/-)	-- --	-- --
total	192(111/81)	6.89 (7.14/6.54)	88 (57/31)	5.43 (5.46/5.37)

Figure 21-a: Breakdown of INT by education for the Penampang Kadazan and Rungus studies

Key:

N = number of cases (men/women)
EDUC = years of formal education
INT = average of individual IT scores (men/women)



(Figure 21-b continued over)

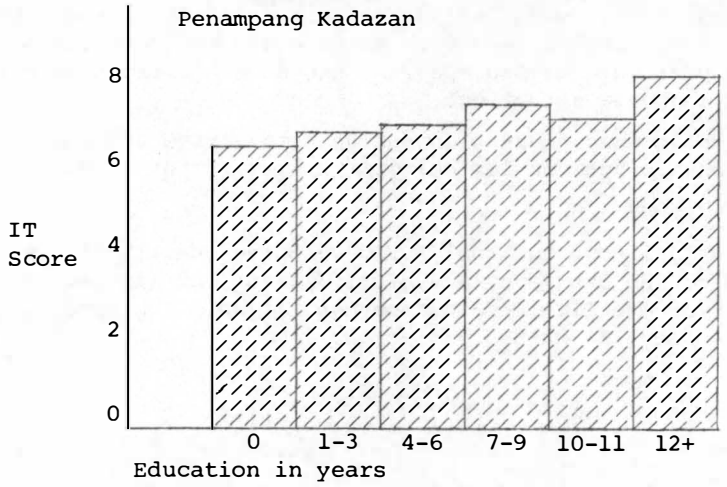


Figure 21-b: IT score vs. education for Rungus and Penampang Kadazan studies

3.3 Summary

We have presented a detailed comparison of two superficially similar language situations, Rungus and Penampang Kadazan. It is clear that a very different set of criteria would be used to predict a person's ability to understand Rungus from that which would be used to predict comprehension of Kadazan. In both cases, geographical distance is the single most important factor, and the strength of the correlation between distance and IT score is remarkably similar in the two studies (see Figure 16 above).

Aside from distance, however, no single factor is found to correlate highly with intelligibility in both of the studies. The striking differences between the two cases give us some insight into the complexity of the problem of developing a general predictive model for intelligibility. A fairly sophisticated model would be needed to account for the Penampang and Rungus situations, to say nothing of the hundred or so other major dialects in Sabah. Such a model would probably have reference to lexical similarity, distance, and other factors not measured in the Sabah survey such as language use, language contact and language attitudes.

4. OBSTACLES TO THE DEVELOPMENT OF A PREDICTIVE MODEL

Simon (1969) defines *secondary analysis* (or "data dredging") as "searching for new relationships in existing data", trying to shed light on new problems with data originally collected for some other purpose. Sections 2 and 3 of this paper present a secondary analysis of the Sabah survey data, a statistical analysis aimed at discovering relationships between IT scores and various other factors.

Developing the kind of predictive model for intelligibility that Simons (1979) proposes will necessarily involve secondary analysis of linguistic and sociolinguistic survey data. No one has ever (to my knowledge) designed and

carried out a dialect intelligibility survey solely for the purpose of developing a general model of intelligibility. Such surveys are so expensive and exhausting that they are only carried out when specific pieces of information are needed about a particular language situation.

Simons (1979:ch.5) used regression analysis to derive a formula expressing the general relationship between lexical similarity and intelligibility:[10]

INT = 1.67 LEX - 66.7

Regression analysis is a powerful tool for developing quantitative models, but when it is applied to the data from a typical dialect intelligibility survey, three sources of error are likely to be present: 1) sampling bias; 2) non-normal distribution of the data; 3) masked variation.

4.1 Sampling error

When a researcher tries to generalise any observed pattern from a particular data sample to an entire population, it is crucial that the sample be fair[11] and representative. The normal way to ensure fairness is to use some form of random sampling; and if the sample size is big enough, a random sample is also very likely to be a good representative of the population as a whole.

Unfortunately, no meaningful language survey could be designed based purely on random sampling. The nature of the information required and the complexity of both the data-gathering and the interpretation stages of the task force us to make careful, principled selections at each step of the survey design.[12]

In a study concerned with estimating a particular measurement (e.g. national language comprehension) for an entire population, the sampling procedure is fairly straightforward - or as straightforward as any social research can be. However, in investigating the general relationship between LEX and INT, the sampling problem is several orders of magnitude more complicated. The researcher must choose a set of subjects from a set of villages representing a set of dialect groups, and for each subject choose a set of measurements (i.e. a set of dialects for which the subject's comprehension will be tested).[13]

In terms of survey design, the goals of the survey determine how choices are made at each level. But in terms of statistical analysis or developing a general model, each "principled decision" becomes a possible source of sampling bias.

On the other hand, Simon (1969:263) points out that random sampling is unnecessary for some kinds of research. Many biological traits, for example, are so homogeneous throughout an entire population that any sample at all is adequate; measurement over even a small, non-random sample can be generalised over the entire population.

We have already mentioned the difference between inherent intelligibility and bilingualism in this regard. If X and Y are related dialects, and group X can be assumed to have had no exposure to dialect Y, then every normal adult native speaker of X should have roughly the same ability to understand Y. A relatively small sample is adequate for measuring inherent intelligibility, but this is emphatically not true for measuring bilingualism.

The relationship between linguistic similarity and intelligibility can be thought of as determined by the innate language faculty common to all humans. In studying this relationship, a small, non-random sample is adequate if the IT

scores reflect only inherent intelligibility, or if cases of bilingualism can be reliably excluded, as Simons (1979) attempted to do. Unfortunately, most dialect surveys include mixed intelligibility and bilingualism situations, i.e. related dialects whose members have significant amounts of contact with each others' speech varieties. This seems to be the normal pattern in most areas of Sabah. In such cases, the validity of generalising any observed relationships depends heavily on the reliability of sampling methods used in collecting the data.

4.2 Distribution of the data

Regression analysis assumes that the data are more or less normally distributed. In section 2.2, we showed that the data from the Sabah survey violate this assumption, especially in the distribution of IT scores. The effect of this skewing is to create a regression line that does not appear to fit the data, and to reduce the correlation coefficient, r. The measured correlation is further reduced by the truncated range of LEX values.

Both the skewing of INT values and the truncated range of LEX are characteristic of most Casad-type dialect intelligibility surveys. Intelligibility is rarely tested where it is not expected to exist, and most surveys have focused on the cognate range of 60-90 percent.

Experience in Sabah has shown that it is very difficult to measure low levels of intelligibility accurately. Subjects began to lose interest in a test when they could not follow the story easily, and some people refused to listen to stories they felt they could not understand.

It is probably easier to use some written form of intelligibility testing, rather than tests of oral comprehension, to measure the low end of the INT scale. However, use of written testing materials introduces an obvious sampling bias by selecting only literate subjects. The distribution of INT values could be made closer to normal by varying the level of difficulty of the texts and questions so that only someone approaching native speaker fluency would be expected to score 100 percent, while 50 percent would correspond roughly to the threshold between language and dialect distinctions. However, increasing difficulty of oral tests also reduces the useful range of the test in terms of LEX. Subjects who were willing to listen to the easy test stories used in Sabah, even when the test dialect was only 65 percent cognate with their own, would be less willing to listen to harder stories in that same dialect.

4.3 Averaging

In most survey reports, the average of the raw IT scores for a given village or dialect group is used as the index of that group's ability to understand some other dialect. The individual scores are neither reported nor (in most cases) used in analysing the data.

As mentioned above, the mean IT score is a valid index of inherent intelligibility, though not of bilingualism. Even so, when mean scores are used in correlation or regression analysis, the results are less accurate than they would be if raw (i.e. individual) IT scores were used.

Correlation and regression analysis are based on calculations of the amount of variation within the data set. When mean scores are used, the variation within each test set of 10 subjects from a particular village is masked. If

variations within each test set are too extreme, the correlation and regression results based on mean scores will be meaningless. The potential magnitude of the error increases with the size of the data set (total number of cases).

With only 10 subjects per test, and raw scores ranging from 0 to 10, the amount of possible variation within each test set is fairly tightly constrained. However, with a data set as massive as that from the Sabah survey, this is still a potentially significant source of error.

4.4 Summary

Regression analysis is potentially the best tool for developing descriptive and predictive models for intelligibility. The results are valid and generalisable for IT scores which reflect only inherent intelligibility.

For studies involving bilingualism or mixed intelligibility and bilingualism, the validity of the results will depend on the reliability of the sampling method. No language survey can be based on a purely random sampling procedure, but studies of bilingualism in particular language situations will be more reliable than broad-scale analyses of bilingualism in general, because the sampling problems are much more manageable.

In studying the relationship between LEX and INT, the accuracy of analysis will be reduced if the distribution of either variable differs greatly from the normal distribution. Survey design plays a crucial part in shaping the distribution of the data collected, and traditional dialect intelligibility surveys seem especially prone to skewing the distribution of IT scores toward the higher values. New methods of measuring intelligibility need to be developed to reach an adequate range of LEX while producing approximately normal distributions of both LEX and INT, without introducing new sources of sampling bias (e.g. relying on literate subjects).

Future analysis of intelligibility data should work directly with raw (individual) scores, rather than average or aggregate scores. This approach is planned for further research, now in progress, using the Sabah data.

All three of the problems discussed above become more serious as the data set becomes larger. Perhaps this is why the cumulative effect was so noticeable for the massive Sabah survey data set. The most promising way to minimise these problems may be to adopt the approach of Simons (1979), i.e. by comparing the results of many relatively small studies. At any rate, more studies are needed concerning the nature of intelligibility and the various factors which affect it in specific language situations.

Acknowledgements

Among the many people who have helped me in various ways with this project, special thanks are due to: Chiang On Khiong and Alfred Chan, of the Sabah State Computer Service Unit, Lim Suk Lee, Tim Friberg, Sue Harris, Priscilla Jones, David Moody, and Gary Simons.

NOTES

1. It is well known that the same factors which promote language learning, such as language contact and positive language attitudes, also promote lexical borrowing over longer time periods. So linguistic similarity is not, strictly speaking, independent of the sociological component of intelligibility. However, the effects of this dependence are assumed to be very small in comparison with other factors involved.

2. The skewness measurement for INT (excluding hometown scores) was -0.526; for LEX, it was -0.276. The kurtosis value for INT was -0.504; for LEX, 1.145. For normally distributed data, skewness = 0, kurtosis = 1.0.

3. In some areas of the Philippines, loss of intelligibility is reported even in the 95% cognate range, due to different use of grammatical markers and other particles (Chuck Walton, personal communication).

4. Simons (1979) suggests 40%.

5. In this study, significance figures are computed using a two-tail test, because no prior assumption is made about the direction (plus or minus) of the correlation.

6. The z-test is used to confirm that the difference between men's and women's scores is statistically significant. For the Penampang scores, the z value is 2.05, indicating that the difference is significant at a confidence level of .02. The difference in the Rungus scores is not significant.

7. In the Kadazan sample, men averaged 2.62 on the travel scale (N = 109, std. dev. = 1.043) while women averaged 2.37 (N = 81, std.dev. = 1.089); z = 1.59, meaning that the difference has a significance of .06.

   For the Rungus sample, men averaged 2.53 (N = 55, std.dev. = 1.331) and women averaged 1.87 (N = 31, std.dev. = 0.763); z = 2.92, indicating a significance of less than .002.

8. z = 1.38, sig = .09.

9. For the Penampang study, the difference between farmers and government employees was significant only at the 0.13 confidence level. The difference between government employees and all other groups was significant at the .04 level.

   For the Rungus study, the significance of the difference between farmers and government employees was below .04; between government employees and others, below .02.

10. One of the strengths of Simons' study was that the raw data came from surveys using a variety of different methods for testing intelligibility. Thus, while INT would be defined operationally as "average score on an intelligibility test", the testing method is not specified. The implied claim seems to be that the relationship expressed in the formula is independent of the testing method used.

11. "Fair" in this sense means that each individual in the population has an equal chance of being included in the sample.

12. Choosing the sample for the Sabah survey involved several levels of decisions. The first question was, at which villages should data be collected? The basic goal was to get data from at least one village from every dialect

group that was either reported (by local residents) or observed (by survey technicians) to be distinctive; and to get a geographically representative sampling of villages from the larger dialect groups. When possible, linguistically homogeneous villages were chosen; but since this tends to be the norm in Sabah, it was not a major constraint. Other factors considered included reported "purity" of language, reported or observed prestige factors, migration patterns (preference being given to long term residents of an area, rather than recent arrivals from other language areas), accessibility and the results of previous survey work by other scholars. The advice and guidance of local government officials was crucial in these decisions, particularly in the first phase of the survey (collection of wordlists and texts).

For each village where intelligibility testing was done, the second question was: which tapes (i.e. which dialects) should be tested here? As discussed in 2.1 above, lexical similarity and distance were primary considerations in determining which dialects should be tested with each other. A further consideration was the desire for comparability between tests. As much as possible, one good test was used to represent a particular dialect everywhere that dialect was tested (rather than a random choice among the tapes recorded in that dialect). For example, the tape from Kampung Bunduon, Penampang, was used in all the Penampang Kadazan testing discussed in sections 2 and 3.

The third level of sampling was the selection of ten individuals to take the test in each village. The strategy called for a rough quota based on age and sex - some old men, some young men, some old women, some young women (no-one under 15). Within these guidelines, the village headmen were generally responsible for finding the subjects.

Each of the factors listed above is a possible source of sampling bias - although some factors could tend to offset each other, e.g. accessibility and prestige vs. "purity" of language.

13. One possible strategy would be a stratified sampling of tests for each subject in the sample group, based on cognate percentages with the subject's own dialect. For a subject from group X, we would divide all other dialects in the state into five sets: 1) all dialects 80-99 percent cognate with X; 2) 70-79 percent cognate with X; 3) 60-69 percent cognate with X; 4) 50-59 percent cognate with X; 5) below 50 percent cognate. The subject would be tested with one dialect selected at random from each group. Needless to say, a survey of this type would be a logistical nightmare, and the results would be virtually useless for any other purpose, such as determining linguistic boundaries or mapping patterns of communication.

REFERENCES

CASAD, Eugene H.

1974 *Dialect intelligibility testing.* Summer Institute of Linguistics Publications in Linguistics and Related Fields, Number 38. Norman, Oklahoma: Summer Institute of Linguistics of the University of Oklahoma.

COLLIER, Ken J.

1977 Predicting intelligibility: a suggested technique. In Loving and Simons, eds, 1977:253-262.

KING, Julie K. and John Wayne KING, eds

1984 *Languages of Sabah: a survey report*. *PL*, C-78.

KIRK, Roger E.

1978 *Introductory statistics*. Monterey, CA: Brooks/Cole.

LOVING, Richard and Gary F. SIMONS, eds

1977 *Language variation and survey techniques*. *Workpapers in Papua New Guinea Languages*, vol.21. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

MILLER, Carolyn P.

1984 Introduction. In King and King, eds, 1984:51-57.

SANKOFF, Gillian

1969 Mutual intelligibility, bilingualism and linguistic boundaries. *International days of sociolinguistics*, 839-848. Rome: Istituto Luigi Sturzo.

SIMON, Julian L.

1969 *Basic research methods in social science*. New York: Random House.

SIMONS, Gary F.

1977 Tables of significance for lexicostatistics. In Loving and Simons, eds, 1977:75-106.

1979 *Language variation and limits to communication*. Technical Report No.3. Ithaca, New York: Department of Modern Languages and Linguistics, Cornell University.

SMITH, Kenneth D.

1984 The languages of Sabah: a tentative lexicostatistical classification. In King and King, eds, 1984:1-49.

WALTER, Steve and Steve ECHERD

n.d. Interpreting intelligibility asymmetries in dialect survey data. MS.

SOME HISTORICAL LINGUISTIC CONTRIBUTIONS TO SOCIOLINGUISTICS

R. David Zorc

0. INTRODUCTORY COMMENTS[1]

Historical linguists can have some pretty strange and varied bedfellows. During the past 16 years of work in the comparative linguistic area, I've needed recourse to such diverse fields as oceanography, botany, entomology, geography, anthropology, and sociology. Forays into distinctly non-Austronesian camps, such as Sanskrit, Chinese, Spanish, and Dutch, have also been necessary, as well as the other 'camps' of theoretical and socio-linguistics. The latter has been particularly helpful and productive, since the way people of different sex, age, and social standing speak profoundly affects the course of language change. Li (1980, 1982a, 1982b), for example, has made Atayalic forms more comparable to Austronesian etyma by unravelling female conservatism from what may be termed 'male speech disguise'.

When it comes to paying long-accumulated debts, most benefactors have to accept simple gratitude. But historical linguists can repay sociolinguists with some insights into determining the exact linguistic situation of multilingual countries and areas like the Philippines, Indonesia, Melanesia, and Oceania. Although I will be discussing the Philippines in particular, what I have to say should apply (certainly *in principle*) to a wide range of language families where the number and the relationships of speech varieties are in dispute. The 100-meaning list presented below is intended to offer a tool for 'fine-tuning' on linguistically discrete communalects*. It is anticipated that a far larger number of such speech varieties will surface than anyone has previously surmised.

1. LANGUAGES, DIALECTS, OR COMMUNALECTS?

In contrast with some popular (non-technical) points of view,[2] a linguist determines a *LANGUAGE* on the basis of mutual intelligibility, whether total (*L-simplex*) or chained (*L-complex*) (see Hockett 1958:327f). Thus, every speech variety is a *DIALECT*, and the combination of all dialects that can communicate directly or indirectly with one another makes up a single *LANGUAGE*. Further refinements have been made, recognising the speech of a single individual as an *IDIOLECT*, and that of a reasonably homogeneous social group as a *COMMUNALECT* (or *ISOLECT* (Hudson 1967)).

In the Philippines alone, there are probably *50 million idiolects* (based on a 1984 population estimate) broken up into approximately *5000 communalects* (based on the number of barrios, sitios, or barangays in non-metropolitan areas),

---

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 341-355. *Pacific Linguistics*, C-93, 1986.

i.e. where people talk in much the same way. While these numbers are very high (and hence not entirely informative), they are a matter of fact. Anyone who has journeyed from town to town within a purportedly common linguistic area (be it Bikol, Panay, Ilocos, or Mountain Province) can attest to the multiple differences in pronunciation, intonation, vocabulary, and grammar prevailing. The situation is certainly the same in Sabah, Java, Sumatra, and other Austronesian locales. In all fairness to the speakers who so choose to identify themselves on the basis of even minor language variations, linguists and laymen alike should accept the communalect as the bottom line. We therefore recognise, as do the speakers, a Marinduque vs Batangas Tagalog, an Oas vs Polangui Bikol, a Kalibonhon vs Libakawnon Aklanon, an Ilianen vs Livunganen Manobo, a Tina vs Botolan Sambal, an Amganad vs Kiangan Ifugao, and so on. Language labels such as 'Bisaya', 'Ifugao', 'Manobo' in this context are uninformative and confusing.[3] Some have been downright wrong, such as 'Sinauna Tagalog' (which is a distinct Southern Luzon language) or 'Datagnon Mangyan' (which is a West Bisayan dialect with no special genetic affiliation to any Mangyan language).

2. THE DETERMINATION OF COMMUNALECTS - A TOOL

The precise number of communalects can be determined by a survey of 100 (or even 50) words that in statistical terms have a high probability of replacement, or, conversely, a low probability of retention (see Dyen, James and Cole 1967). Table 1 is derived from principles discussed a decade ago (Zorc 1974) and virtually separates the Philippines into several thousand linguistic communities. For a positive score (+) in this kind of comparison, it is essential that forms be identical in sound, accent, form, and grammar - any difference whatsoever is crucial in establishing a communalect, and hence should be counted as minus. While historical linguists and lexicostatisticians are concerned with cognates (forms descended from a single ancestral word or etymon), sociolinguists take notice of differences separating speakers.

In scoring this list, for example, Tagalog laró? differs from Sinauna lalú? *play* (r vs l), and each differs from Alangan ladó?, even if all three descend from an etymon *ladú?. Tagalog (um)akyát, Kapampangan mukyat, Aborlan-Tagbanwa apyat, and Ivatan k(um)ayat *climb* again differ from one another, even though they are ultimately cognate. Phonological differences (Kinaray-a bédlay :: Hiligaynon búdlay *difficult*; Aklanon ?índi? :: Tagalog hindí? *not*; Romblon huyát :: Aklanon hu+át *wait*), accent dissimilarities (Bontok ?ótot :: Pangasinan ?otót *rat*), and semantic mismatches (Tagalog do?ón *there-far* :: Northern-Samar du?ún *there-near*) need to be regarded as separators of communalects.

Table 1: Differential vocabulary separating Philippine communalects

	TAGALOG	CEBUANO	AKLANON	NAGA BIKOL	ILOKANO	W.BKD MANOBO	ILIANEN MANOBO	CEN.MIND. MANOBO
01.*afternoon	hápon	hápun	hápun	hápun	malém	hapun	məʔapun	
02.*angry	galít	sukú?	ʔákig	ʔaŋgút	ʔuŋét	{paʔuk / ʔəpəs	laŋət	{kəpaʔuk / kələŋətan
03.*ashamed	hiyá?	ʔúlaw	huyáʔ	súpug	baʔín	ʔələd	yəyaʔan	
04.*bad/evil	masamá?	dáʔut	maɬáʔin	maraʔút	dákes	daʔat	məraʔat	məraʔat
05. bark (tree)	balát	pánit	pánit	ʔúpak	ʔukís	ləkaŋ	ʔupis	
06. beautiful	magandá	níndut	mayadʔáyad	magayún	napintás	{dagwəy / tandaŋ		tandaŋ
07.*bird	ʔíbon	láŋgam	píspis	gamgám	billít	tagbis	pəpənuk	tagbis
08. blow (at)	hípip	huyúp	huyúp	hayúp	puyút	hiyup	pəriyup	pəriyup
09.*boil (intr.)	kuló?	bukál	bukáɬ	kalaʔkágaʔ	burék	diʔdiʔ	diʔdiʔ	
10. bright	maliwánag	lámdag	maháyag	liwánag	naranyág			
11. carry/bring	dalá	dalá	daɬáh	dará	yeg	ʔuwit		baba
12.*chest	dibdíb	dúghan	dúghan	daghán	barúkuŋ	kumən	kumən	kumən
13.*chin	bába?	suwáŋ	suɬáŋ	kuʔkúʔ	tímid	baka?	baka?	baka?
14.*climb (up)	ʔakyát	saká	sáka?	sakát	ʔumúli	{təkəzəg / pəmənahik	ʔambak	{təkərəg / ʔambak
15.*cold	{magináw / malamíg	túgnaw	maɬamíg	malíput	{lamʔek / lamíʔis	gənəw	məʔadsil	məʔadsil
16. collapse	gibá?	gubá?	gubá?	gabá?	narbá			
17. command	ʔútos	súgu?	súgu?	súgu?	bílin	{sugu? / təlaʔan		təlaʔan
18. companion	kasáma	ʔubán	kaʔibáhan	kaʔibá	kaduá	duma	duma	duma
19. crawl	gápaŋ	kámaŋ	kámaŋ	kamáŋ	karayám	pənanap		
20. crush-lice	tirís	ʔirúk	turús	tadús	ligʔís			
21. cut-off	pútol	pútul	putúɬ	putúl	púted	raprap		
22. dark	madilím	ŋítŋit	maduɬúm	diklúm	nasipŋét	mərusirəm		mərusirəm
23. different	ʔibá	láʔin	ɬaʔín	{ʔibá / láʔin	sabáli	{saŋi? / sələkəw	laʔin	səŋəkuʔən
24.*difficult	mahírap	lisúd	malisúd	dipísil	narígat	mərəgən	mərəhən	mərəhən
25. dirty	marumí	húgaw	hígku?	maʔatíʔ	narugít	rəmərik	mərədsik	mərədsik
26.*dust	ʔalikabók	ʔabúg	taputapúh	ʔalpúg	tápuk	ʔəliyavuk	lipukpuk	ʔalyabuk
27.*earth/soil	lúpa?	yúta?	ɬúgta?	dagáʔ	dagá	tana?	tana?	tana?
28. fall (down)	húlog	húlug	húɬug	húlug	tinnág	{ʔulug / piləy	piləy	ʔulug
29.*fast/swift	mabilís	páspas	páspas	kaskás	nadarás	gaʔan	məsasəw	
30.*few	kaunti?	gamáy	saŋkurút	diʔít	bassít	dəʔisəy		
31.*fight/quarrel	ʔáway	ʔáway	ʔáway	ʔíwal	riŋgúr	{ʔəgət / təbək		{ʔəhət / kəwahaʔən
32.*finger	daliri?	túdluʔ	túdluʔ	murúʔ	rámay	{kəmər / təzuʔ	təruʔ	kəmər

Table 1 (cont'd)

	TAGALOG	CEBUANO	AKLANON	NAGA BIKOL	ILOKANO	W.BKD MANOBO	ILIANEN MANOBO	CEN.MIND MANOBO
33. *fish*	ʔisdáʔ	ʔísdaʔ	ʔísdaʔ	siráʔ	ʔikán	paʔit	səraʔ	{seraʔ ŋalap
34. *flood*	bahá?	bahá?	bahá?	bahá?	layús			
35.*forget	límot	límut	lipát	liŋáw	límpat	lipat	lipat	
36.*good	mabúti	maʔáyu	mayád	maráy	nasayáʔat	ʔupiya	məʔupiya	
37. *happy*	{masayá maligáya	lípay	malípay	maʔugmá	naragsák	məlipəy		galəwhaləw
38. *hard* (subst)	matigás	gahíʔ	matígʔa	matagás	nataŋkán	məzəsən	mərəsən	mərəsən
39.*here	díto	{dirí díʔnhi	{ʔiyá diyá	digdí	ditúy	{kayi dini	kayi	kay
40.*hold	háwak	kupút	buyút	kapút	ʔiggém	{gawəd kəmkəm	kəmkəm	{gawəd kəmkəm
41. *kind*	mabaʔít	buʔután	mabúʔut	mabúʔut	naʔánus	tulaŋəd		tulaŋəd
42. *lake*	{láwaʔ lánaw	línaw	danáw	dánaw	dánaw	ranəw	ranəw	ranaw
43.*lie/untruth	sinuŋáliŋ	bakák	purí l	pútik	ʔulbúd	taruʔ	taruʔ	taruʔ
44. *lonely*	mapaŋláw	{míŋaw guʔúl	namíŋaw	mapuʔŋaw	nalidáy	buluŋ	kəlimərawan	
45. *lose/lost*	walá?	walá?	dúɬaʔ	waráʔ	púkaw	tazin	tarin	tarin
46.*many	marámi	dághan	ʔabúʔ	dakúl	ʔadú	məzakəl	mərakəl	mərakəl
47.*morning	ʔumága	búntag	ʔagáhun	ʔága	bigát	məsələm	məsələm	məsələm
48. *mud*	pútik	lápuk	ɬúnaŋ	labúy	pítak	basak	basak	basak
49.*narrow	{makítid masikíp	síɡpit	makitíd	kipút	ʔakíkid	məligət	məlihət	məlihət
50.*near	malápit	duʔúl	maɬapít	haranî	ʔasidág	{ʔuvəy rani	mərani	ʔubəy
51. *noisy*	maʔíŋay	sábaʔ	masáŋag	marîbuk	naʔariyaŋgá	məzagiŋ		
52.*none/nothing	walá?	walá?	ʔuwáʔ	mayúʔ	ʔawán	waza?	wara?	wəraʔ
53.*not (fut.)	hindíʔ	díli?	ʔíndiʔ	daʔí	saʔán	kəna?	kəna?	kəna?
54. *one (as in counting)	ʔisá	ʔusá	ʔisaɬáh	sarúʔ	maysá	{ʔisa səvəha	səbəka	{ʔisa səbəka
55.*play	laróʔ	dúlaʔ	hámpaŋ	káwat	ʔayʔáyam	baləyvaləy	dərəmət	galəw
56. *push*	túlak	tulúd	tuɬúd	ʔúsul	túlak	{tukuʔ tulud	sinumag	{sumag tulud
57. *put/place*	lagáy	butáŋ	butáŋ	bugták	kábil	savuk		tahu
58. *question/ask*	tanóŋ	paŋutána	paŋutána	hapút	saludsúd	ʔinsaʔ		ʔinsaʔ
59. *rainbow*	bahagháriʔ	baláŋaw	baɬaŋáw	{baláŋaw bulalaŋáw	bullaláyaw	{bəludtu bəlugtu	bəluntu	bəluntu
60. *raincloud*	dagím	dágʔum	gáɬʔum	daʔgúm	libuyúŋ	kivəl	kibəl	kibəl
61.*rat	dagáʔ	ʔilagáʔ	ɬáŋgam	kinúʔ	baʔú	rumat	rumat	{kiput rumat

Table 1 (cont'd)

	TAGALOG	CEBUANO	AKLANON	NAGA BIKOL	ILOKANO	W.BKD MANOBO	ILIANEN MANOBO	CEN.MIND. MANOBO
62.*river	ʔílog	subáʔ	subáʔ	sálug	karayán	wahig	wayig	wayig
63. round	mabílog	líŋin	malibúnug	talímun	nabukél	kəlizəŋ		
64. sad	maluŋkót	subúʔ	masubúʔ	mamundúʔ	naladíŋit			
65. say/said	sábi	súlti	hámbaɬ	sábi	saʔú	kaɡi	kahi	kahi
66. seek	hánap	paŋítaʔ	ʔúsuy	hánap	bíruk			
67.*short (obj)	maʔiklíʔ	mubúʔ	matágʔud	haliʔput	ʔababá	məvavaʔ	məbaba	məbabaʔ
68. short (pers)	pandák	putút	putút	hababáʔ	pandék	{pəndak / məlimpuguʔ	məbaba	məbabaʔ
69.*sibling	kapatíd	ʔigsúʔun	ʔigmáŋhud	túgaŋ	kabsát	suləd		
70.*sit	ʔupóʔ	líŋkud	líŋkud	túkaw	tuɡáw	pinuʔu	pinuʔu	pinuʔu
71.*slice (meat)	híwaʔ	híwaʔ	kíwaʔ	pidásu	ʔíwa	karad	karad	karad
72. slow	mabágal	hínay	mahínay	malúway	nabuntúg	{məlanat / nanəy		nanay
73.*small (obj)	maliʔít	gamáy	maʔisút	sadáy	bassít	dəʔisək	dəʔisək	dəʔisak
74. smell (n.)	ʔamóy	báhuʔ	húgum	páruŋ	ʔáŋut			
74a.smell (v.)	ʔamuyín	timahúʔ	hugúm	parúŋun	ʔaŋúten	hazək	ʔəbpəŋarək	ʔarək
74b.fragrant	mabaŋó	humút	mahumút	mahamút	nabaŋlú	həmut	məʔəmut	məʔamut
74c.bad-smelling	mabáho?	bahúʔ	mabáhuʔ	mabatáʔ	nabuyúk	məmahuʔ		məmawuʔ
75. soft	malambót	humúk	mahumúk	malumúy	naluknéŋ	ləminəg		{mələmək / məʔuməl
76. space under the house	síloŋ	síluŋ	síluŋ	síruŋ	síruk	sihuŋ	siyuŋ	
77. speak/talk = word	salitáʔ	púluŋ	hámbaɬ	tarám	saʔú	lalag	lalag	
78. spider (gen)	{gagambá / ʔanlaláwaʔ}	lawaʔlawáʔ	ɬamáŋ	láwaʔ	lawalawá	kələlawaʔ	kəlawaʔlawaʔ	
79. split = cleft	biʔák	buʔák	bukáʔ	buwáŋ	{bísak / búsak}	tevíʔ		
80. tear/rip	púnit	gísiʔ	gísiʔ	gísiʔ	pígis	bindas		
81.*that (far)	ʔiyón	kádtu	datú	ʔitú	daydiáy	həʔazaʔ	ʔəyan	ʔayan
82.*there (far)	doʔón	didtu	{ʔídtu / ditú}	dumán	didiáy	diyaʔ	kənyan	
83.*this	ʔitó	{kirí / kiní}	dáya	ʔiní	daytúy	həʔini	ʔini	ʔini
84.*throat	lalamúnan	tutúnlan	tutúnlan	halanúhan	karabukúb	bəkərəŋ	bəkərəŋan	bəkərəŋan
85.*throw	hágis	lábay	habúy	ʔapún	ibatú	{timbag / ʔantug}	ʔantug	timbag

Table 1 (cont'd)

	TAGALOG	CEBUANO	AKLANON	NAGA BIKOL	ILOKANO	W.BKD. MANOBO	ILIANEN MANOBO	CEN.MIND. MANOBO
86. **throw-away*	tápon	lábug	pilák	tápuk	bellé[ŋ]	timbag	timbag	timbag
87. **tomorrow*	búkas	ʔúgmaʔ	hinʔágah	sa ʔága	ʔintun bigát	kəʔəsələm	kəʔəsələm	kəʔəsələm
88. *turn (in a direction)*	likóʔ	líkuʔ	likúʔ	síkuʔ	sikkú			tiku
89. *turn/revolve*	ʔíkot	túyuk	tíyug	tarírik	pusípus	biyu	tilə[ŋ]	
90. *ugly*	páŋit	ŋilʔad	maɬáwʔay	makanús	nalá?ad			
91. *vagina*	púki	{bilát bútuʔ	{bilát puyás	{budáy putáy	ʔúki	bətiʔ	bətiʔ	
92. **wait*	{hintáy ʔantáy	hulát	huɬát	halát	ʔúray	tagad	tahad	
93. *waterfall*	tálon	busáy	busáy	busáy	dissúʔur	ʔəvaga	dəmpilas	sampəw
94. **weak*	mahína?	lúya	maɬúya	malúya	kapsút	{məguyaʔ məluya	məlubəy	
95. *wear/put-on*	suʔót	súlʔub	súksuk	suʔlút	ikapét	sunʔud		
96. **west*	kanlúran	kasádpan	katuɬúndan	{subsúban sulnúpan	láʔud	sənləpan	sənləpan	sənləpan
97. **what?*	ʔanú	ʔúnsa	{ʔanúh nánuh	ʔanú	ʔanya	həŋkəy	ʔəŋkəy	ʔəŋkəy
98. **when?* (fut.)	kaʔilán	sánʔu	hinʔunúh	nuʔarín	kaʔánu	kəʔənu	kəʔənu	kəʔənu
99. **wide*	malápad	lapád	maɬápad	halakbáŋ	ʔakába	məluʔag	məluʔag	
100. *wrong*	malíʔ	sayúp	saɬáʔ	salá?	(dákes)	salaʔ		

When gathering data for a survey of this sort, it is imperative that exact semantic matches be obtained. While the list is designed to exaggerate differences amongst even close dialects, any cause(s) for such separation should be real and not the result of inexact comparison. The following notes are included to guide researchers as to the semantics intended.

00 - All forms elicited should follow the matching of English and Tagalog. I have not been able to double-check the data with informants, but I have noted some errors in the main sources (McFarland 1977 for Tag, Ceb, Naga, Ilk; Reid 1971 for WBM and Iln) when compared to other published data (Elkins 1968 for WBM; Vanoverbergh 1956 for Ilk; Wolff 1972 for Ceb; Mintz 1971 for Naga). Akl is from Zorc 1969, and CMM from Elkins 1954.

02 - The most general term for *anger*; avoid: *peeved, upset, crabby.*

03 - Here and throughout the list, verb affixes have not been included. If affixes are included, a single paradigm should be obtained, e.g. *I was embarrassed* (simple past); affixes could be used to show differences beyond the root word or stem.

04 - The most common term for *bad* often coincides with the word for *evil*. A sense such as Tag masamá? aŋ panahón *the weather is bad* or masamá? aŋ pakiramdám ko *I feel bad* is intended.

05 - Often the same as the form for *skin* (which is omitted from the list). Avoid specialised terms for *second layer of bark* (CMM luwit) or *bark/skin of banana trunk* (Akl ?úpas).

06 - Usually the opposite of *ugly* (#90); *good looking* as applied to a woman, especially if *handsome (male)* is differentiated in the language.

08 - As in *blowing at/on a fire to increase its heat*; avoid: *blowing out (as a match)* or *exhaling.*

09 - The actual boiling of water (intransitive verb); avoid: *to boil (vegetables/eggs)* (Tag lágа?) or *inception of boiling when first bubbles are formed,* etc.

10 - Generic; avoid: *brightness of sun or moon, glare.*

11 - The most general form for carrying or bringing something from one place to another regardless of means of transport or carriage. Avoid: *carry on back* (Tag babá), *carry in the hand* (Tag bitbít), *carry on shoulder* (Tag pasán), *carry on head* (Tag súnoŋ), etc.

12 - The upper torso, not just *breast* or *ribs.*

13 - Distinguish from *jaw* (Tag síhaŋ, paŋá).

15 - As in *cold weather* (Tag maginάw, Ilk lam?ék); distinguish from *cold (to touch)* (Tag malamíg, Ilk lamí?is); either could be used in the comparison, so long as the sense is the same.

16 - As a house from age or a ship from a storm.

17 - As from a person in authority; not just *send on an errand,* nor *deathbed command = final will and testament* (Tag bílin).

19 - As a baby on all fours; distinguish from *crawl on one's belly* or *crawl as a snake* (WBM dula).

20 - Crush between the fingernails.

21 - As in amputating a limb.

23 - Several languages distinguish between *another - of the same type* (Akl ʔibáh) vs *another - of a different kind* (Akl ɬaʔín); the latter sense is intended here.

25 - Avoid special senses such as *dirty laundry* (Akl ʔumúg), *muddy*, etc.

27 - Not: *earth/world* (Tag daʔigdíg) or special kinds of soil, e.g. *clay*.

30 - A little bit, in small amount.

31 - Not: *hit, box, come to blows*; more in the sense *to fall out*.

32 - Generic for *digit* (often the same as for *toe*); avoid: *thumb, index finger, middle finger*, etc.

36 - Generic; avoid senses such as *good at* (Tag magalíŋ) or *clever* (Tag marúnoŋ).

40 - As in: *hold this for me*; avoid specifics such as *hold in the palm of one's hand* or *hold under the arm* (see note #11).

41 - As a good or giving person; avoid *loving, generous*.

43 - As in telling a deliberate untruth; avoid senses such as Tag magbuláʔan *fib, exaggerate, 'bullshit'*.

44 - Different from *sad* (#64); sense of isolation or melancholy felt.

45 - Note that only verb affixes differentiate this from *none* (#52) in Tag and some other CPh languages; differentiate from *lose one's way* (Akl táɬaŋ).

49 - As a narrow entrance or road; Tag masikíp *tight, crowded* is perhaps too specialised.

52 - *There is none*.

53 - The future negative, as in *I will not go*. Note the various negators in McFarland 1977:20; only one of five possibilities is sought here.

54 - In the sequence: *one - two - three* .... Note that numerous forms with classifiers occur (e.g. Akl saŋka-, sambáto, sambílog).

56 - Generic; not specialised meanings (e.g. WBM dəgupi *be pushed along or aside by a strong force*, dəkuzas *push something back and forth over a surface*, pəsəl *push something with the thumb or finger*, etc.).

60 - As distinguished from other types of cloud.

65 - Avoid quotative particles (e.g. Tag daw, Akl kunúh).

66 - To look for something that has been *lost* (#45).

69 - Generic; avoid terms for *older* and *younger* sibling.

70 - Generic; avoid terms for *squat, sit on the ground*, etc.

71 - Not just *cut*; avoid specifics like *to slice thin* or *to slice into big chunks*.

72 - Generic adjective; avoid verbs like *to cook slowly* (WBM nanəy *to proceed slowly* may be too specific here and is wrongly included).

73 - Distinguish from *small amount = few* (#30); note that Ilokano (and perhaps other communalects) do not draw such a distinction.

74 - 74a-c are included here to illustrate the various senses that can exist within a given meaning and the dangers of semantic mismatch. The most

neutral *noun* for *smell* is intended (with no implications as to the pleasantness, or otherwise, of the aroma), not verbs (74a) or adjectives (74b,c).

75 - The opposite of *hard* (#38); not that of *loud, coarse*, etc.

77 - May not be differentiated in some communalects from *say* (#65) (e.g. Aklanon).

79 - As a stone or wood may split open; distinguish from senses such as Tag bisák *to split wood*, Tag háti? *to split in half* (= Akl píhak).

80 - As cloth or paper.

85 - Keep separate from senses such as *throw stone(s), throw overhand, throw underhand, throw-away* (#86); generic *to throw* [*x*] *at, cast.*

88 - Generic; avoid verbs like *turn right, turn left, turn one's head.*

89 - Intransitive verb, like a coin or top turns around.

95 - As in *"What will you wear?"*, avoiding specific garment terms.

96 - Sometimes not distinguished from *sunset-place* or *west-wind.*

98 - As in *"When will he arrive?"*; some languages have *when* (past)?.

99 - Opposite of *narrow* (#49).

100 - *In error, incorrect*; not intended as a negator (Akl bukún *not so*).

This list may be used in toto, or those 50 items marked with an asterisk may form an abbreviated survey. If the criteria outlined above are strictly applied, only those speech varieties that score 90% (in excess of 45/50 or 90/100) with one another can be regarded as belonging to the same communalect - and if the speakers consider themselves as such. In this way, there is a sociological and linguistic confirmation of a given (Philippine or Austronesian) language scene. Because of borrowing, common inheritance, and convergence (e.g. disparate shift of *p > f, *d > r, *r > l, *ə > u, etc.), scores will rarely be 0. Ilokano, for example, scores 4% with Akl and Ceb, up to 8% with Tag. However, the list has been constructed on the basis of abundant data (Reid 1971, Yap 1977, McFarland 1977), so that it can be stated with confidence that scores will be very low, even between reasonably close genetic relatives.

The languages chosen in Table 1 demonstrate this last point. Tag, Ceb, Akl, and Naga are genetically related Central Philippine languages. Akl and Ceb, which are Bisayan, score no higher than 42%; Tag-Bik, no higher than 21% with each other. Central Mindanao Manobo stands in a dialect relationship with Ilianen and Western Bukidnon, yet the CMM-Iln score is 88% and CMM-WBM is 76%, while Iln-WBM is 63%. Even if cognates, rather than identities, are counted, the Manobo scores are: CMM-Iln 91%, CMM-WBM 88%, and Iln-WBM 76% *based on the data available* (numerous lacunae for CMM and Iln make these rough computations). Thus, Central Mindanao Manobo (or Kiriyenteken Manobo; Elkins, personal communication 9 August 1983) is a communalect in its own right.

Of just the 50 items marked with an asterisk, Ilokano has 31 unique forms, Tagalog 23, and Cebuano 11 (the latter is due to Ceb's strong influence in the central and southern Philippines resulting in numerous loans into or from Ceb). These uniques dictate that no other communalect could share a score higher than 19/50 with Ilk, 27/50 with Tag, or 39/50 with Ceb, except a communalect that was indeed Ilk, Tag, or Ceb respectively. In fact, Tagalog scores 8/50 with Sinauna, 7 with Kapampangan, 4 with Botolan, and 3 with Bikol. Cebuano scores 25/50 with Hiligaynon (due to loans), 18 with Samar-Leyte, 12 with Surigao, and 10 with

Aklanon. Ilokano scores 11/50 with Itneg, 8 with Kankanay, and 7 with Luba. The overall effects of convergence are thus negligible in this kind of survey, which is *sociolinguistic* and *not historical* in intent.

3. THE DETERMINATION OF LANGUAGES

Although we may eventually know how many communalects exist in the Philippines or other Austronesian areas (since adequate data are available in the files of the Summer Institute of Linguistics and several researchers), if we address the question of how many languages there are, numerous problems beset us. Since a language is defined in terms of mutual intelligibility, both the degree and the kind of intelligibility would need to be determined (see Zorc 1977:165-170). Some linguists would accept *genetic intelligibility*: if a Malay says "Mata ku sakit" (*my eye hurts*), and if a Filipino understands him (as most would be likely to), then obviously some communication is taking place. But the Malay may rattle on and virtually all of the rest of the speech act could well be lost on the Filipino. This is not *practical intelligibility* - the Malay is not likely to get much joy from a Filipino doctor if each sticks to his own language. The Summer Institute of Linguistics needs to know the degree to which a translation of the Bible can be understood by speakers/readers in other areas. They have conducted extensive tests of intelligibility throughout the Philippines; if too many barriers to understanding exist, a different translation is necessary. Each speech variety is accorded its own dignity; linguistic imperialism is avoided - Warays may understand Cebuano or Aklanons Hiligaynon, but each deserves their own intimate version.

If linguists could agree on a criterion for determining mutual intelligibility (the SIL tests and scores are accurate and sound in this regard), and factors such as bilingualism and sesquilingualism[4] (when someone understands but cannot speak another language) could be controlled, then we would be well on our way to knowing how many dialects and languages there are in the Philippines. The exact answer could be known within this decade, depending on research interests of M.A. or Ph.D. scholars and access to SIL files. While SIL has always been most generous and open with its data, it would be most appropriate (given the years of labour and research involved) if an SIL member drew up a comprehensive Philippine matrix of intelligibility test scores, possibly as part of his/her studies for a degree. Kroeger's paper on "Intelligibility patterns in Sabah" in this volume is a welcome step in this direction.

In the meantime, genetic linguistics can provide a working solution. The number of languages in the Philippines alone has been debated and estimated by linguists and laymen. Blumentritt (1901) recognised 194 native groups mentioned in the literature of his time. But he well knew that many of these were repetitious or inaccurate in several ways. Conklin (1952), being more concerned with linguistic criteria, outlined 75 main groups broken up into a total of 156 members. Historical/comparative linguists are generally in agreement that there are, at most, 28 major linguistic groups that can be described as 'Philippine' on the basis of geographic or genetic criteria (see Table 2). One subgroup, Sama (11), is clearly intrusive to the Philippines within the last millenium, and is genetically of an 'Indonesian type', possibly related to the South Sulawesi group (including Makassarese and Buginese (Roger F. Mills, personal communication, 4 October 1983)).

Table 2: Probable and possible Philippine subgroups

N1 North Cordilleran
- 1 South (Gaddang-Yogad)
- 2 North (Atta-Ibanag)
- 3 Central (Malaweg, Isneg)
- 4 Agta

N2 Dumagat = East Cordilleran
- 1 Negrito
- 2 Paranan
- 3 Central (Casiguran)
- 4 South (Umirey)

N3 Ilokano

N4 Central Cordilleran
- 1 South (Isinai)
- 2 North (Itneg)
- 3 East (Kalinga)
- 4 Nuclear (Balangaw)
- 5 Ifugao
- 6 Bontok-Kankanay

N5 Ilongot

N6 South Cordilleran
- 1 Pangasinan
- 2 Inibaloi-Karaw
- 3 Kallahan

N7 Bashiic = Ivatanic
- 1 Yami
- 2 Itbayaten
- 3 Ivatanen-Babuyan

N8 Southern Luzon
- 1 Sambalic
- 2 Sinauna
- 3 Kapampangan

N9 North Mangyan
- 1 Iraya
- 2 Alangan
- 3 Tadyawan

N10 Inati of Panay

S1 South Mangyan
- 1 Buhid-Taubuid
- 2 Hanunoo

S2 Palawanic
- 1 North (Aborlan-Batak)
- 2 South (Molbog-Brookes)

S3 Kalamianic

S4 Central Philippine
- 1 Tagalog
- 2 Bikol
  - -Inland
  - -Coastal
  - -Pandan
- 3 Bisayan
  - -West
  - -Banton
  - -Central
  - -Cebuan
  - -South
- 4 Mansakan
  - -Mamanwa
  - -North
  - -East
  - -West

S5 Subanon
- 1 Siocon-Kalibugan
- 2 Sindangan-Salug-Lapuyan

S6 Manobo
- 1 North
- 2 Inland
- 3 South

S7 Danao

S8 Bilic
- 1 Giangan/Bagobo
- 2 Tiruray
- 3 Tboli
- 4 Inner Blaan

S9 Sangiric
- 1 North (Sangil-Sangir-Talaud)
- 2 South (Bantik-Ratahan)

C1 Minahasan
- 1 South (Tonsawang)
- 2 North (Tontemboan)
- 3 North-east (Tondano-Tonsea-Tombulu)

C2 Mongondow-Ponosakan

C3 Gorontalic
- 1 Dila (Buol-Suwawa)
- 2 East (Bulanga)
- 3 West (Gorontalo)

I1 Sama-Bajaw
- 1 Indonesian Bajaw
- 2 North Borneo/Sabah Land Bajaw
- 3 Jama-Mapun
- 4 Southern Sulu
- 5 Central Sulu
- 6 Western Sulu
- 7 Northern Sulu
- 8 Yakan
- 9 Zamboanga Sama
- 10 Abaknun

U1 Chamorro

U2 Palau

U3 Yapese

B1 Kadazan-Dusunic

B2 Murutic

Code: B = Borneo
C = Celebes
N = Northern Philippines
S = Southern Philippines
U = Ungrouped

Maps showing the locales of these languages can be found in McFarland 1980.

Within the geopolitical boundaries of the Philippines, there are at most 19 language groups (N1-10 + S1-9) which could share an immediate genetic ancestor, which have in all likelihood developed in situ over at least 3,000 years, and which can not be attributed to multiple migrations from overseas as popular history suggests. Note, however, that three groups are represented in northern Celebes (Sulawesi, C1-3) which can be proven to be immediately related to Southern Philippine languages. There are two families in Borneo (B1-2) which share features of Philippine and the distinct Sabahan languages. Blust (1974) has proposed that these are more intimately related to the North Sarawak subgroup, based on the sharing of the innovation involving strengthened reflexes of PAN *b, *d, *j, *g. Their similarities to the Ph-type may be due to loans or convergence, and require further study. Another three (U1-3) are spoken in the Pacific and await definitive classification. As linguistic research progresses, these groups will probably be collapsed, but the current state of knowledge and debate dictates some prudence, so that the maximum number (19 Philippine + 1 Indonesian intrusive) represents a core of agreement amongst Philippinologists, amidst otherwise widespread disagreement as to the collapsibility of these to ten (Ruhlen (in progress)), or two, or even one. (See my paper on "The genetic relationships of Philippine languages", where I argue for the latter alternative, i.e. a common Proto-Philippine ancestor from which all Ph languages except Sama descend.)

While we can be sure that there are no more than *20 major linguistic groups* (N1-10, S1-9, I1) within the Philippines, speakers would take little consolation in such broad criteria. Cebuanos identify themselves as Bisaya (not Central Philippine); the same holds true for Bikolanos or Tagalogs; and, more widely, for Ibanags, Pampangeños, etc. Hence, Table 2 delineates *50 Northern and Southern Philippine subgroups* with which speakers may more readily identify.

The verification of these as languages (based on the bounds of mutual intelligibility) and their fragmentation into communalects (recognising dialectal idiosyncrasies) must await further study.

What was it that I was saying earlier about gratitude? I have just outlined a massive task - for Philippinologists and for Austronesianists! I have presented a 'fine-tuning tool' for isolating communalects and given some suggestions from the genetic evidence for what ultimate language groups we may arrive at. The hard work ahead may not be appreciated, but hopefully the hints will be helpful.

NOTES

1. Some of the ideas discussed in this paper, including the original 50 items from Table 1, have appeared in Zorc 1984. The present paper and its FOCAL companion (Zorc 1986) split the topics covered therein, and treat them in much greater detail. I am grateful to Otto Dahl for a list of five Malagasy dialects which confirms the value of Table 1 in differentiating communalects, and to Paul Black for many helpful comments on the original draft.

2. Many Filipinos regard a *language* as a widespread and prestigious vehicle of communication (such as English, Chinese, Russian, or Pilipino), while any other kind of speech is 'a *dialect*'. This view is compatible in many regards with the concept of communalect discussed below. In practice, Filipinos are aware of even the most minute linguistic variations and label them accordingly (even if not always complimentarily, e.g. "They talk like birds").

3. Because they refer to genetic subgroups (if a linguist is talking), or geopolitical isolates (if a layman is talking).

4. I once met some Tadyawan-Mangyans who claimed they could understand Ilokano! The genetic gap (by *any* measure) between Ilokano and Tadyawan is so great that the only cause for such a statement was their frequent dealings with an Ilokano merchant (who bought bundles of rattan from them). Intelligibility must be tested by rather precise (rather than impressionistic) measures, if it is going to be validly established. It is for this reason that I expressed reservations (Zorc 1977:170 and footnote 59) about links between Sorsogon :: Bikol and Naturalis :: Kamayo. The lexical and grammatical differences between these speech varieties must create a considerable amount of *code noise* (Hockett 1958:331f) and render intelligibility far from perfect and mutual. I rather suspect sesquilingualism (or passive language ability) has led to such statements. If, however, *all* members of *both* communities are sesquilingual, then a link genuinely exists. But in a certain area of Davao City, all the people on a block understand Tagalog, Hiligaynon, and Ilokano. An Iloko will speak Ilokano to an Ilonggo, who will respond in Ilonggo. Would this mean that Ilokano is now part of the Bisayan complex? One would (hopefully) not seriously propose this!

*EDITORIAL NOTE: Zorc (1984 and elsewhere) uses the spelling communilect, and not the more usual communalect, as used by *Pacific Linguistics* here.

REFERENCES

BLUMENTRITT, Ferdinand

1901 Philippine tribes and languages. *Annual Report of the Board of Regents of the Smithsonian Institution*, 527-547. Washington.

BLUST, Robert A

1974 The Proto-North Sarawak vowel deletion hypothesis. Ph.D. dissertation, University of Hawaii.

CONKLIN, Harold C.

1952 Outline gazeteer of native Philippine ethnic and linguistic groups. Reprinted in Andrew B. Gonzalez, Teodoro Llamzon and Fe Otanes, eds *Readings in Philippine linguistics*. Manila: Linguistic Society of the Philippines, 1973.

DYEN, Isidore, A.T. JAMES and J.W.L. COLE

1967 Language divergence and estimated word retention rate. *Language* 43:150-171.

ELKINS, Richard

1968 *Manobo-English dictionary*. *Oceanic Linguistics* Special Publication No.3. Honolulu: University of Hawaii Press.

ELKINS, Richard and Betty ELKINS

1954 *A vocabulary of Central Mindanao Manobo*. Manila: Summer Institute of Linguistics, Bureau of Public Schools, Institute of National Language.

HOCKETT, Charles F.

1958 *A course in modern linguistics*. New York: Macmillan.

HUDSON, Alfred B.

1967 *The Barito isolects of Borneo*. Data Paper Number 68. Ithaca: Southeast Asia Program, Department of Asian Studies, Cornell University.

KROEGER, Paul

1986 Intelligibility patterns in Sabah. In this volume, 309-339.

LI, Paul Jen-kuei

1980 Men's and women's speech in Mayrinax. In *Papers in honor of Prof. Lin Yü-k'eng on her seventieth birthday*, 9-17. Taipei: Academia Sinica and National Taiwan University; Wen Shin.

1982a Male and female forms of speech in Atayal. *The Bulletin of the Institute of History and Philology* 53/2:265-304. Taipei: Academia Sinica.

1982b Linguistic variations of different age groups in the Atayalic dialects. In *Studies in linguistics presented to Dr. Fang-kuei Li on his eightieth birthday*, 167-191. *Tsing Hua Journal of Chinese Studies*, New Series 14. Taipei.

McFARLAND, Curtis D.

1977 *Northern Philippine linguistic geography*. Study of Languages and Cultures of Asia and Africa Monograph Series No.9. Tokyo.

1980 *A linguistic atlas of the Philippines*. Study of Languages and Cultures of Asia and Africa Monograph Series No.15. Tokyo.

MINTZ, Malcolm W.

1971 *Bikol dictionary*. PALI Language Texts: Philippines. Honolulu: University of Hawaii Press.

PALLESEN, A. Kemp

1977 Culture contact and language convergence. Ph.D. dissertation, University of California at Berkeley.

PENNOYER, F. Douglas

n.d. Unpublished field notes on the Inati dialects of Panay (Inete, Sugodnin 1, Sugodnin 2).

REID, Lawrence A.

1971 *Philippine minor languages: word lists and phonologies. Oceanic Linguistics* Special Publication No.8. Honolulu: University of Hawaii Press.

RUHLEN, Merritt

in progress *A guide to the world's languages*.

VANOVERBERGH, Morice, C.I.C.M.

1956 *Iloko-English dictionary*. Baguio City: Catholic School Press.

WOLFF, John U.

1972 *A dictionary of Cebuano Visayan. Philippine Journal of Linguistics,* Special Monograph No.4. Manila: Linguistic Society of the Philippines.

YAP, Fe Aldave

1977 *A comparative study of Philippine lexicons.* Manila: Institute of National Language, Department of Education and Culture.

ZORC, R. David

1969 *A study of the Aklanon dialect,* vol.2: *Aklanon to English dictionary.* Kalibo: Aklan Printing Center.

1974 Towards a definitive Philippine wordlist - the qualitative use of vocabulary in identifying and classifying languages. FICAL paper. *Oceanic Linguistics* 13/1&2:409-455.

1977 *The Bisayan dialects of the Philippines: subgrouping and reconstruction. PL,* C-44.

1984 The Philippine language scene - the sociolinguistic contribution of historical linguistics. In Andrew Gonzalez, FSC, ed. *Panagani essays in honor of Bonifacio P. Sibayan on his sixty-eighth birthday,* 135-148. Manila: Linguistic Society of the Philippines.

1986 The genetic relationships of Philippine languages. In *FOCAL II,* 147-173.

A COMPUTATIONAL APPROACH TO STRESS PATTERNS IN PENRHYN

Ayako Yasuda-Graefe and Volker Graefe

1. PURPOSE

'Stress' or 'accent' in the Polynesian languages has been described as predictable, dependent on penultimate-and-alternate-preceding-vowels rules and first-vowel-and-long-vowel-precedence rules (Biggs 1961; Hohepa 1967; Newbrand 1951). However, many exceptions to such rules have also been reported. This is also the case in Penrhyn, a Polynesian language spoken in one of the Northern Cook Islands (Yasuda 1968).

The first difficulty when dealing with the phenomenon of stress is to identify its nature objectively. In this study we use the term 'stress' to indicate the prominence of a syllable relative to other syllables in a stretch of speech (Folkins et al 1975). It is one of our goals to find an objective, quantitative measure for the perceived stress. Such a measure would be valuable in two respects: It would facilitate the formulation and the testing of theories or rules related to stress patterns, and it would be the first step towards the utilisation of stress to improve the performance of speech understanding machines.

It is also our goal to demonstrate that all the necessary experiments for such studies can efficiently be conducted using no other tool than an ordinary digital computer with a few simple peripheral devices to input and output speech signals.

A third goal is to find out to what extent modelling the psychoacoustic properties of the human auditory sense can contribute to a better understanding of the nature of human speech and to the development of better speech-understanding machines.

2. METHOD

What we perceive as emphasised or 'prominent' is related to other perceptual qualities such as pitch, loudness and tempo of speech segments, whose acoustic correlates in many languages have been said to be changes in fundamental frequency, intensity and syllable duration respectively (e.g. Lehiste 1970; Folkins et al 1975). It is therefore necessary to extract these acoustic parameters from the speech signal.

A great variety of instruments exists for the analysis of speech signals. Among them are sound spectrographs, level meters, filters, fundamental frequency meters and many more. Using these instruments correctly is not always easy.

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics,* 357-376. *Pacific Linguistics*, C-93, 1986.

Some of them are quite complicated and if the operator does not fully understand their internal working and their inherent limitations, gross errors may occur. All these instruments tend to age and to give unreliable results if they are not properly and diligently maintained. Moreover, if the same recording of a speech signal is analysed with different instruments it can be difficult to establish the exact temporal correspondence between their outputs. (Does the peak in fundamental frequency precede or follow the peak in intensity?)

We have avoided all these difficulties by using only one piece of equipment, a digital computer. A suitably programmed digital computer can perform all the functions of the various instruments mentioned above and many more. Therefore our speech samples were stored in digital form in the computer and all analysis was performed using the methods of digital signal processing.

There are great advantages to this technique, especially in respect of precision, reproducibility and ease of use. A digital filter, for instance, will never drift nor age; if so designed it will not introduce any phase lag, nor noise, nor distortion; its passband will be flat and its rolloff will be steep. All of its characteristics can easily and predictably be modified just by changing a few lines of program code. The same holds true for all other methods and tools of digital signal processing.

2.1 Equipment

Figure 1 shows the equipment we used for our experiments. The speech samples were available as tape recordings and an ordinary tape recorder was used for playback.

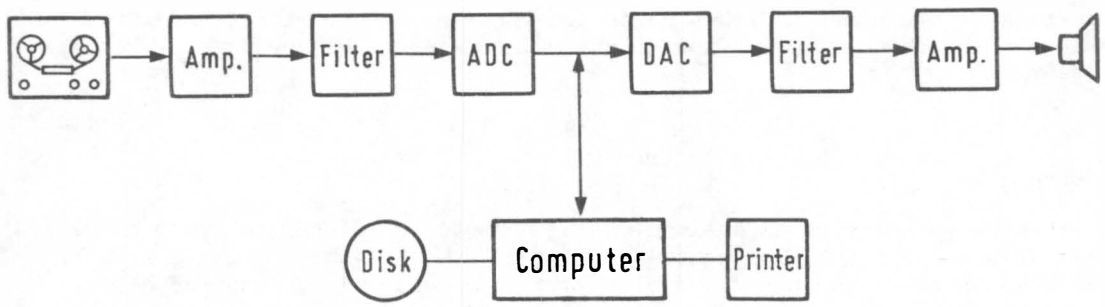


Figure 1: The equipment used for the analysis of the speech data

To store speech samples in the computer, the output from the tape recorder was passed through a pre-amplifier and a bandpass filter (70 Hz - 10 kHz) to an analog-to-digital converter (ADC) with 12 bits of resolution and a sample rate of 25 kHz. The digitised speech signal was fed into the computer, and simultaneously into a digital-to-analog converter (DAC), also with 12 bits of resolution. The reconstructed analog signal was again bandpass-filtered (70 Hz - 10 kHz), amplified, and fed into a loudspeaker to permit the continuous monitoring of the digitised signal.

The same setup, with the ADC switched off, was used to listen to speech samples after they had been stored in the computer.

The computer is a Perkin Elmer 3210 with 2 MB of memory and 160 MB of mass storage capacity. A matrix printer (Facit 4542) with limited graphics capabilities was used to produce all hard copy output, including curves and spectra.

As figure 1 shows, the hardware we have been using is very simple, as all the complexity is hidden within the computer and its software. Because of its simplicity it is reliable and easy to operate; since the equipment involved is always the same, regardless of the type of analysis being performed at any given moment, nothing ever has to be moved around or reconfigured.

The equipment did not introduce any audible noise or distortion; therefore we did not find it necessary to conduct specific measurements to evaluate its performance quantitatively.

2.2 Data

The speech samples were recorded during a field study in 1966-67 in Auckland, New Zealand. A small consumer-type battery operated tape recorder with a primitive microphone was used to record the utterances of the informants. Although several informants from Penrhyn supplied data, the main informant was an elderly woman, who spoke only Penrhyn and Rarotongan. The data were analysed grammatically in detail by Yasuda (1968).

Unfortunately the recordings are of poor technical quality in several respects:

(a) Noise

The dynamic range is only about 35 dB. Not only tape hiss is present but also motor noise from the tape recorder, traffic and household noise (the recording was done in the informant's living room) and sometimes even voices in the background.

(b) Frequency range

Practically no signal is present above 3 kHz.

(c) Nonlinear distortions

Occasionally the tape recorder was overloaded in spite of its automatic volume control.

These deficiencies are severe. The high noise level limits our studies to the louder parts of each utterance, while the limited frequency range limits us mostly to the voiced parts.

On the other hand, for many practical applications it is important to develop the ability to analyse speech signals which are band limited and corrupted by noise. Also, we are interested in stress patterns, and these stress patterns are clearly audible even in our low-quality tape recordings.

2.3 Programs

2.3.1 Speech input/output

All input and output of speech signals is performed by one interactive program. It includes the following functions:

- Transfer of selected speech samples from the tape recorder into the computer's memory;
- Transfer of speech samples, together with written comments, between storage disk and main memory;
- Playback of stored speech samples (either the entire sample or only selected parts of it).

In playback mode the operator can control the exact start time and end time (relative to the beginning of the speech sample) of the segment he wants to hear. Start time, end time, or both can also be incremented or decremented automatically in operator selected steps. This makes it fairly easy to determine with great accuracy the borders of each audible segment within the speech sample.

2.3.2 Sound pressure level

A typical plot of sound pressure level versus time is shown in figure 2. The level is computed for each 10-ms-interval. Since the actual sound pressure level at the time of the recording is not known, it is assumed that the strongest signals which were recorded on the tape correspond to a sound pressure level of 80 dB.

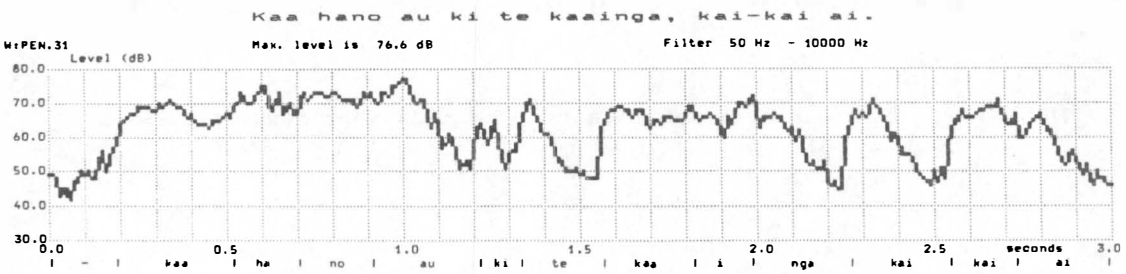


Figure 2: Typical plot of computed sound pressure level versus time

(As in all of the following computer-printed diagrams, the width of the original diagram is about 30cm, and the time scale is relative to the beginning of the digital speech sample.)

It is possible to apply various numerical filters to the data before computing the level. The important advantages of numerical filters compared to physical filters have already been pointed out: all their characteristics are freely selectable and perfectly stable, and they do not introduce any noticeable phase lag, distortion or noise.

2.3.3 Spectrogram

Sound spectrograms have been an important means to analyse speech signals since they were introduced by Koenig et al (1946). Figure 3 shows a sound spectrogram as it has been produced by the computer. Compared to using a conventional sound spectrograph, computing the spectrogram has several advantages.

Temporal and spectral resolution can easily be varied, there are no errors caused by phase lag, noise, or distortion and all parameter settings are perfectly reproducible.

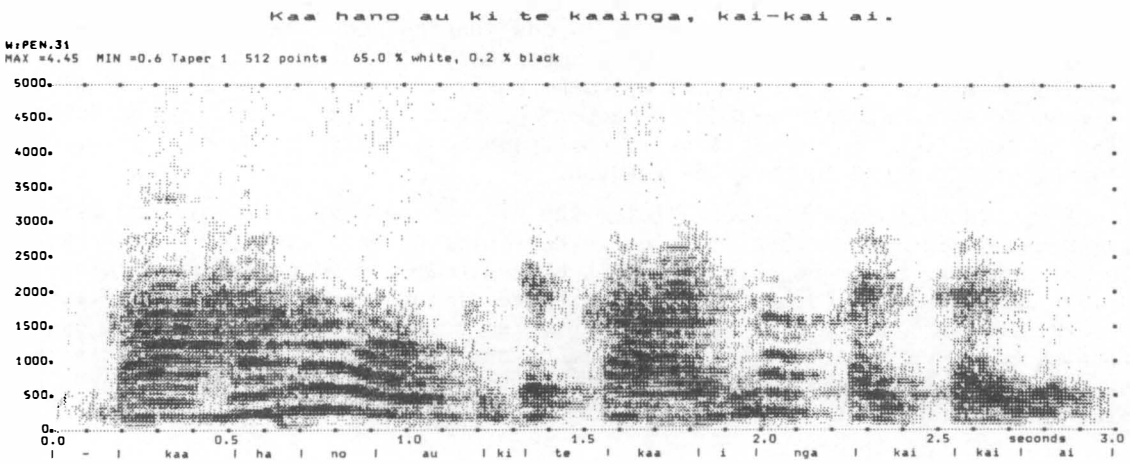


Figure 3: Typical computed sound spectrogram

(The frequency range in this example is 0 to 5000 Hz.)

In figure 3 one spectrum was computed for each 10-ms-interval. Because in speech signals the high frequency components tend to be weaker than the low frequency components, the data were differenced, which is equivalent to using a 6 dB per octave high-pass filter. For each spectrum 512 data points were used (0.02s) and a cosine window was applied before performing the fast Fourier transformation. This results in a spectral resolution of about 100 Hz.

2.3.4 Loudness

All speech signals consist of louder and softer segments, and it is reasonable to expect that the fluctuations of loudness are related to the syntactic or semantic structure of the utterance. Loudness is not a physical quantity but a perceptual or subjective one; two sounds are of equal loudness if the 'average' listener perceives them to be equally loud. Loudness is, of course, related to physical quantities like sound pressure and frequency, but in a very complicated way which is not fully understood yet, in spite of many years of effort by numerous researchers. The basic facts are (Zwicker and Feldtkeller 1967; Zwicker 1982):

- If the frequency of a sound signal is varied while its intensity and other characteristics are kept constant, the loudness varies in a characteristic way. The loudness is greatest in a frequency range around 1 kHz to 4 kHz.
- If the intensity of a sound signal is varied while all other characteristics are kept constant, the loudness doubles whenever the sound pressure level is increased by 10 dB. (Strictly speaking, this is true only if the sound pressure level is at least 40 dB.)

- The auditory system forms critical (frequency-) bands. Below 500 Hz the width of each critical band is approximately 100 Hz, above 500 Hz it is approximately 20% of the centre frequency. If a sound signal has several components within the same critical band, they contribute to the loudness according to the sum of their intensities. If it has several components in different critical bands, components in one critical band can mask components in other bands. The masking is not symmetrical, essentially only components at higher frequencies are masked by components at lower frequencies. If a component is completely masked it cannot be heard and it does not contribute to the perceived loudness; if it is partly masked, it can be heard, but its contribution to the loudness is reduced.

These facts were established in series of psychoacoustic experiments using static sound signals which did not carry any information. Probably the perception of loudness of speech signals which are dynamic and do carry information is even more complicated, but we are not aware of any experimental data related to such signals. Therefore, we wrote a program to compute the loudness of speech signals, based on an algorithm that would be correct for static noise. This is not a completely satisfactory method, but we consider it a first step towards a more adequate model of the auditory system.

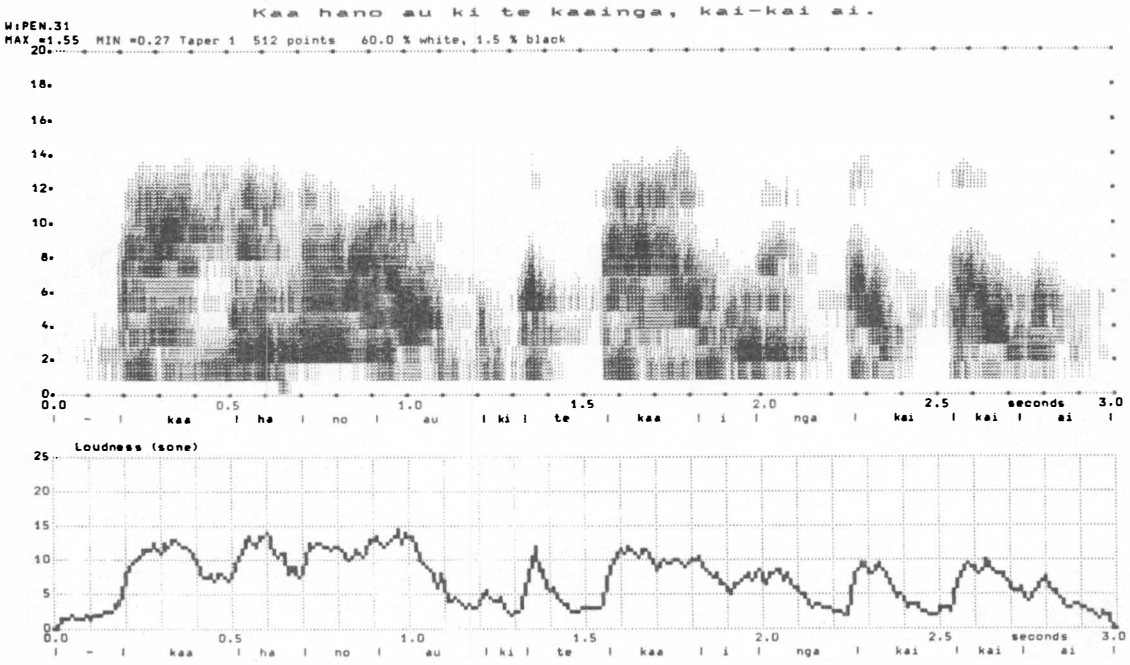


Figure 4: Typical output of the loudness program

(Top: specific loudness versus time and pitch;
pitch is given in units of Bark = 100 mel.
Bottom: computed loudness versus time.)

Figure 4 shows typical results of the program to compute loudness. The program includes the following steps:

- Computation of the power spectrum as described in section 2.3.3;
- Combining adjacent spectral lines into critical bands;
- Computation of loudness and specific loudness, using a subroutine by Paulus and Zwicker (1972). Specific loudness indicates how strongly components of different pitch contribute to the total loudness. Pitch is a perceptual quantity, corresponding closely to the physical quantity frequency.

2.3.5 Fundamental frequency

Several methods are known to determine the fundamental frequency of a speech signal (for an overview see, e.g. Markel and Gray 1976). They all utilise one of two basic ideas: they either evaluate a periodicity in the frequency domain (which also manifests itself as a pattern of horizontal stripes in the sound spectrogram), or they evaluate a periodicity in the time domain directly. Finding the fundamental frequency automatically is difficult, and apparently no method is known which always works reliably.

We have used two methods simultaneously: a sound spectrogram to indicate the general trend of the fundamental frequency within an utterance, and the frequency of zero-crossings to get quantitative data (figure 5). Where the zero-crossing-data are too irregular we use the spectrogram to correct them manually.

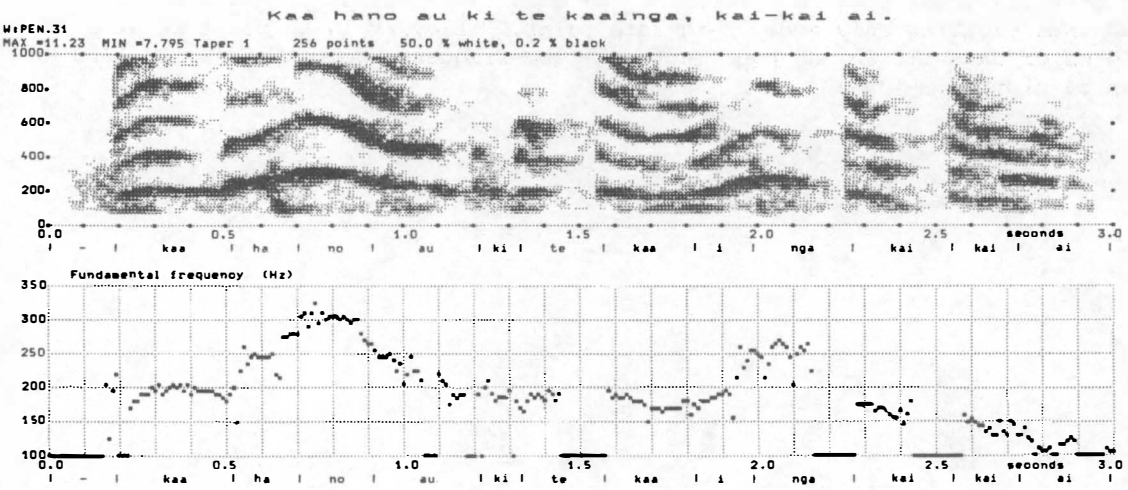


Figure 5: Output of the program for finding the fundamental frequency

(Top: power spectrum, 0 to 1000 Hz.
Bottom: fundamental frequency from zero-crossings.)

Before the zero-crossing rate can be used to compute the fundamental frequency, all components of the speech signal which are outside the fundamental frequency range should be filtered out. In our speech samples the fundamental frequency is almost always between 100 Hz and 300 Hz, therefore we used a filter with a passband from 100 Hz to 300 Hz for most of our speech samples.

A threshold is used to detect pauses and unvoiced segments; ideally a value of zero should in such cases be returned for the fundamental frequency, but because of the poor signal-to-noise-ratio of our speech samples it is difficult to select the correct value for the threshold.

For a fully automatic system it would probably be necessary to use adaptive filters and to adjust the threshold automatically according to the momentary signal level, but for our purpose it was acceptable to use constant settings of the filter and of the threshold, and to correct irregular data manually.

2.3.6 Waveform plot

Sometimes it is useful to scrutinise the data in very great detail, for instance, to analyse some unexpected phenomenon that can be heard, to determine the exact location of syllable borders in situations where even careful listening yields no clear result, or to find out why the fundamental frequency program behaves erratically at a certain point within the speech sample.

For these and similar purposes a program exists which plots the waveform data on paper (figure 6).

The difficulty with such a program is the mass of data, 25 000 values for each second of speech. If every single data point is plotted, 36 ms of data fill one standard page. The plot quickly becomes unreasonably large, unless only very short sections of the signal are plotted. Therefore, in another mode of operation, the program compresses the data by first applying a low-pass filter and then plotting only every n-th data point. If every 10-th point is used, 600 ms of data fit in one page, but only low-frequency components up to 1200 Hz can be plotted.

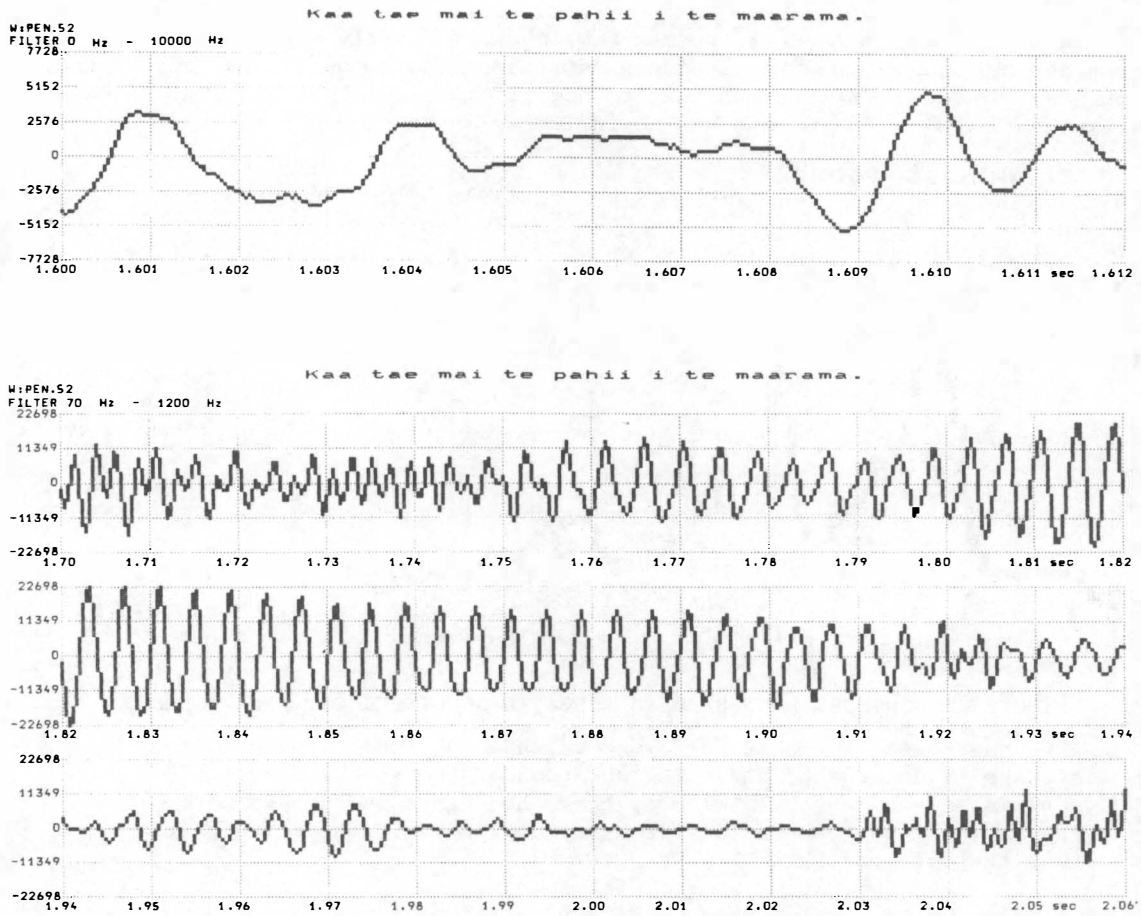


Figure 6: Waveform plots

(In the upper diagram every data point of a 12-ms-section of an utterance has been plotted, in the lower diagram every tenth point of a longer section has been plotted after removing all high frequency components above 1200 Hz.)

3. THE NATURE OF STRESS

In this section we discuss the relationship between perceived stress and the acoustic parameters which we have extracted from the data. In particular, we study loudness, syllable duration and fundamental frequency as possible counterparts to perceived stress. Since stress is primarily an attribute of syllables, we also study the correspondence between acoustic parameters and the syllable structure of utterances.

3.1 Loudness and intensity

In declarative sentences the maximum loudness usually occurs in one of the first few syllables; then the loudness decreases slowly towards the end of the sentence. There are utterances that do not fit into this pattern, but we have no explanation for it. Stress has no influence on the deviation.

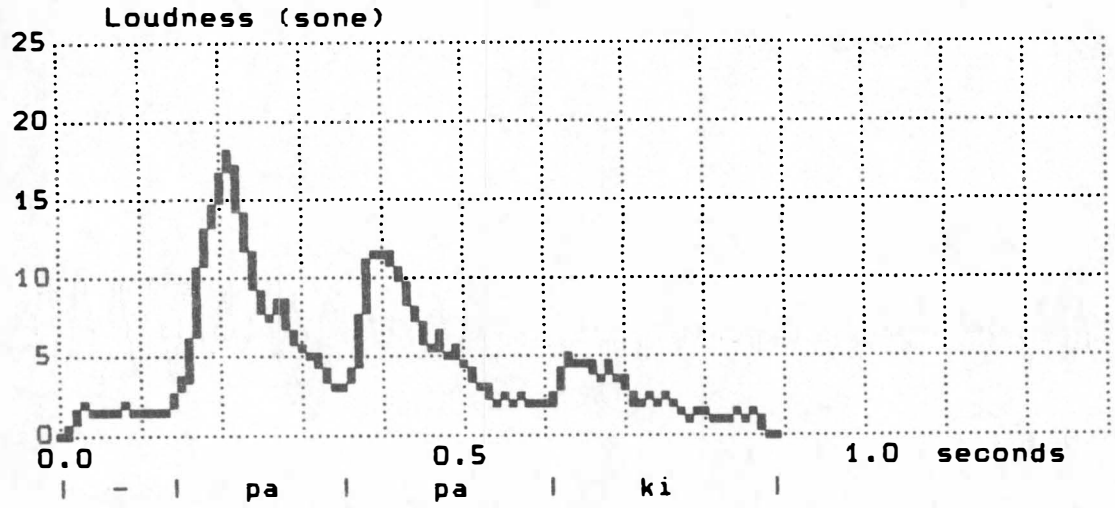


Figure 7: Loudness (measured in sone) in an utterance: papAki *push*

(The unstressed syllable /pa/ is more than 50% louder than the stressed counterpart /pA/. The duration of the stressed syllable /pA/ is approximately 10% longer than the unstressed counterpart /pa/. (cf. the fundamental frequency of /pA/ is, however, higher than that of /pa/.))

At this point we have not found any direct relationship between stress and loudness. It seems, however, that between certain vowels there are intrinsic differences in loudness. [a] is normally higher than [i] in loudness scale, but when [i] is stressed, it is heard a little louder, sometimes almost as loud as [a] in the same stretch of speech.[1] Other vowels do not show any relevant data in terms of stress. As a matter of fact, there are many cases where an unstressed vowel is even significantly louder than its stressed counterpart (figure 7).

The plots of loudness (figure 4) and sound level or intensity (figure 2) are, however, quite useful as a guide for the segmentation of Penrhyn utterances into syllables. Both types of data can be used, but the loudness plot usually has a better structure, clearer appearance and is easier to work with. Very often the nucleus of a syllable coincides with a maximum of loudness and the syllable boundaries with minima. The reason is that vowels tend to be louder than consonants (with the exception of nasals). This is particularly true for voiceless stop consonants, which consist of a short period of silence, followed by a sharp rise in sound level (often 30 dB in 10 ms). Since many Penrhyn syllables begin with a voiceless consonant,[2] a large share of all syllable boundaries can be detected by visual inspection of the loudness diagram; figure 9 shows an example where most syllable boundaries coincide with loudness minima. The final segmentation has to be done by careful listening to sections of the utterance. This time-consuming task is much easier if good candidates for syllable boundaries are already known.

3.2 Syllable duration

The duration of syllables has often been cited to have a high correlation to stress. In many languages, including English, an increase in duration of a given syllable gives an indication that the syllable is more prominent in a word or utterance (Lehiste 1970).

The measurement of syllable durations is not always an easy task. The difficulty is in the determination of the exact location of the syllable boundaries. If a syllable is enclosed within stop consonants the ends of the stops can be taken as very distinct syllable boundaries (see figure 7). The duration of the syllable can then be determined with an uncertainty of less than 10 ms. Examples are:

/papAki/ *push*: [pa] and [pA]
/kOpe/ *hurry* and /koopUU/ *guts*: [kO] and [koo]

In other cases syllable boundaries are not marked precisely by any acoustic parameter and cannot easily be determined by careful listening either. In such cases the duration of a syllable can be uncertain by more than 100 ms. Sometimes a careful examination of the waveform plots (figure 6) can help in finding syllable boundaries.

We have not found any significant durational differences between stressed and unstressed syllables. At most, a stressed syllable may be perhaps 10% longer than its unstressed counterpart. On the other hand, there are cases where an unstressed syllable is even longer than a stressed one. There are, however, great differences between long and short syllables. A long syllable is often two or three times longer than a short syllable. There are no cases in which the duration of a long syllable is less than that of a short syllable. Even an unstressed long syllable is always significantly longer (1.5 to 2 times) than a stressed short syllable. This result is consistent with the fact that long vowels are phonemic in Penrhyn; an increase in duration gives a phonemic difference rather than a stressed impression.

Therefore, we conclude that an increase in duration may be coincidental to stress, but has no direct relationship to it in Penrhyn.

3.3 Fundamental frequency

A typical and idealised pattern of our informant's fundamental frequency movement in a neutral declarative sentence is shown in figure 8. This general pattern of fundamental frequency is more distinct in a longer utterance (e.g. 5-sec-utterance) than in a shorter one (e.g. 1-sec-utterance). This is a slightly modified pattern of what Vaissière (1983) calls a language independent fundamental frequency contour.

The utterance usually begins with an intermediate frequency (about 150 Hz - 200 Hz) and after a few syllables it reaches the highest frequency (about 250 Hz - 300 Hz). From a communications point of view this appears plausible because it lets the hearer find out quickly which frequency range and tempo the speaker is going to use for the following utterance. The fundamental frequency rises and falls almost periodically as a function of time, but with decreasing amplitude and an overall falling tendency. It reaches the lowest frequency when it ceases voicing near the end of the utterance.

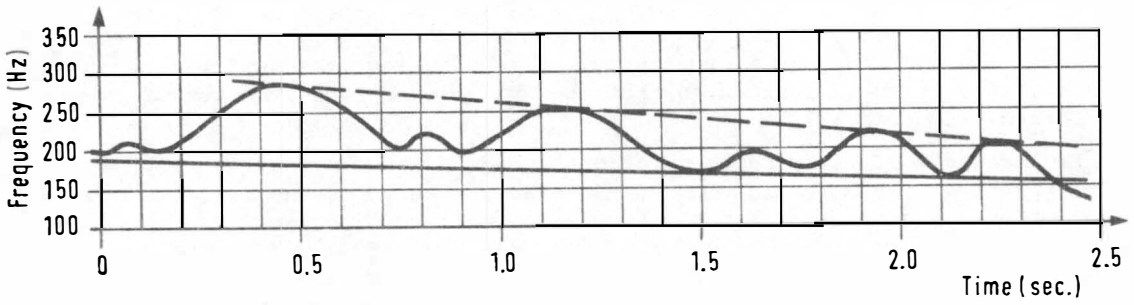


Figure 8: Typical modulation of fundamental frequency in a Penrhyn declarative sentence, forming major waves (touching the upper broken line) and minor waves

(The number of waves is not fixed.)

A very strong relationship exists between perceived stress and the modulations of fundamental frequency. Especially whenever a maximum exceeds the upper broken line, one perceives phonetic prominence very clearly. Usually the stress perceived as 'strongest' in an utterance coincides with such a high peak of the fundamental frequency contour. Besides, along each wave one perceives a stress, too. The closer the peak comes to the upper broken line in figure 8, the more prominent the corresponding syllable appears to be.

In both cases, however, it is not necessarily the syllable which coincides exactly with the peak of the wave but the syllable with the greatest change in fundamental frequency that is perceived as more prominent than others; when the fundamental frequency rises or first rises and then falls immediately or before it falls rapidly, one perceives a phonetic prominence on the corresponding syllable. On the other hand, in an unstressed syllable the fundamental frequency stays near the base line or in the valley of a wave.

From these facts we conclude that stress in Penrhyn is closely related to changes in fundamental frequency. It is not the absolute value of the frequency, but rather its relative value within an utterance that is perceived as stress.

4. STRESS PATTERNS

In general we must take the following points into consideration when we analyse the data:

(a) Stress is an indication of a physical effort of a speaker to make a particular part of an utterance more prominent than others.

(b) Like other suprasegmental features, stress alone does not convey any linguistic information. Stress must always be considered together with segmental and other suprasegmental features.

(c) Stress in carefully spoken slow speech often behaves differently from stress in normal fast speech. Therefore these two styles of speech should be distinguished when dealing with stress patterns (see section 4.2).

(d) A one-word utterance must be considered as an independent short utterance. The stress pattern within such an utterance does not necessarily correspond to 'word stress' but to stress in a short, perhaps carefully spoken utterance.

(e) There are many different types of stress in each language: an intrinsic stress, which all speakers of the language use as part of a natural intonation, a pay-attention stress, which a speaker uses to emphasise a particular information, a stress that expresses emotions, and many others. Although these different types of stress belong to different categories of language analysis, they are encoded using the same set of acoustic parameters.

(f) There is no 'word stress' in Penrhyn that gives a minimal pair such as *pErmit* and *permIt* in English.

(g) In Penrhyn, an intrinsic sentence stress distinguishes a declarative sentence from an interrogative sentence and perhaps from an imperative sentence too. In this paper we only deal with declarative sentences. A superficial look at the data suggests, however, that the stress pattern of a Penrhyn declarative sentence is similar, if not identical to the pattern of interrogative and imperative sentences containing interrogative morphemes and imperative morphemes respectively. Only yes/no questions (interrogative sentences without an interrogative morpheme) differ in stress patterns. However, this has not been tested because of lack of data.

4.1 Syllables

Stress is primarily an attribute of syllables. However, what we mean by 'syllable' is not easy to define. There is no exact universal definition of syllables, although it is a basic building block of many languages.

It has been pointed out in section 3 that syllable boundaries are not marked precisely by any single acoustic parameter. When we use all the parameters (figure 2 to figure 6) and when we listen to the speech data carefully, we may be able to find out a great amount of phonetic change that takes place in a stretch of utterance. However, it is very difficult to segment it, because it is always continuous. It is also extremely difficult to distinguish between consonants and vowels; a certain quality of vowels depends upon the preceding and the following consonant and vice versa.

On the other hand we often discover some significant changes that occur in a number of acoustic parameters. When we use these data, we can determine a minimum phonetic unit and we call it 'syllable'. We define the Penrhyn syllable structure as follows:

> Penrhyn syllables are either long or short. A short syllable has the form [C]V; a long syllable has the form [C]VV. Syllable boundaries occur at every possible breath pause, before every consonant and after every second vowel in a sequence of vowels. In Penrhyn there are no phonemic consonant clusters.

A sequence of two identical vowels does not take any special position in the syllable structure of Penrhyn, because it behaves just like a sequence of two different vowels:

/pooro/ *ball* [pOoro] or [poOro]
/puaka/ *pig* [pUaka] or [puAka]

The stress on the first vowel or the second varies freely from speaker to speaker and from one occasion to another. In slow careful speech the stress occurs more often on the second vowel in both words, and in normal fast speech the stress is likely to occur on the first vowel in both cases.

4.2 Domains of stress

As we mentioned in section 3.3, Penrhyn stress is almost exclusively related to changes in fundamental frequency. In this section we investigate if these changes of fundamental frequency show systematic stress patterns in several domains: sentence, phrase and a certain smaller domain which we call measure.

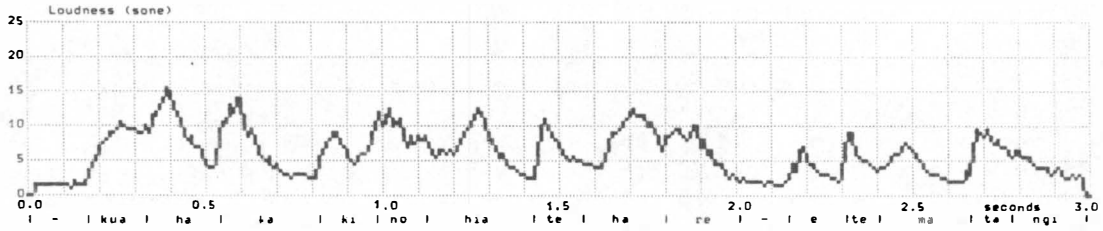


Figure 9: Syllable boundaries and minima in loudness

(#kua haka-kIno-hia te hAre e te matAngi#
The house was damaged by the wind.

The morphologically expected syllables are:
kua ha ka ki no hia te ha re e te ma ta ngi

Most of these syllables are clearly separated by a minimum in loudness.)

4.2.1 Sentence

When we examine the changes of fundamental frequency, we notice that there is often a deviation from the standard contour; a maximum exceeds the upper broken line of figure 8. Here one perceives a very strong stress which usually indicates the most important information of a given sentence. So, we call it 'sentence stress' (figure 11).

The position of a sentence stress is normally semantically determined (see figure 11):

Kaa hano au ki te kaaInga, kai-kai ai.
I am going HOME to have a meal there.

It is interesting to note, however, that a sentence stress occurs very frequently towards the end of an utterance. Because of the structure of Penrhyn sentences, a semantically important morpheme is often placed at the end of a sentence. Also by raising the fundamental frequency at the end, Penrhyn speakers seem to indicate that the utterance is ending. We call this 'sentence-final stress' (see figure 10). A sentence-final stress is usually accompanied by devoicing of the final vowels and decreasing loudness. This may suggest a typical Penrhyn declarative sentence structure, because this 'sentence-final stress' is perceived as 'stress' by a hearer who does not know the language, while it is perceived as 'a natural intonation' by a hearer who speaks or knows the language well.

Some sentences contain no sentence stress (the prosodically neutral sentences), others have one or more, depending upon which information is emphasised by the speaker. The degree of emphasis does not depend on the absolute frequency

value but rather on the relative amount of the deviation: how far the peak exceeds the upper broken line or how much is the difference between the maximum and the preceding or the following minimum.

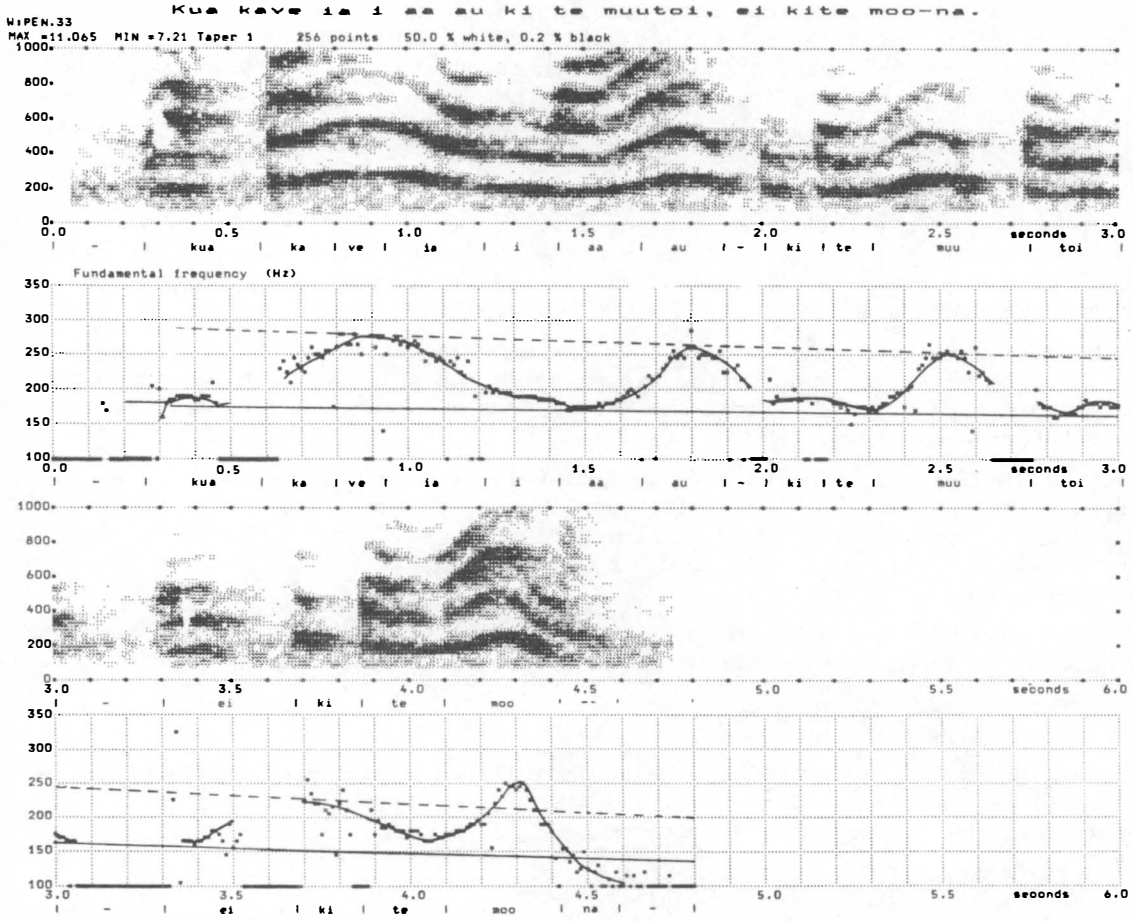


Figure 10: Fundamental frequency of a Penrhyn declarative sentence

(A sentence-final stress occurs at the morpheme m00-na *for him* where a maximum exceeds the upper broken line, and phrase stress occurs along the major waves, which contain lexical morphemes.)

#kua kAve ia i aa AU ki te mUUtoi ei kIte m00-na#
take I police appear
He took me to the police as his witness (to appear FOR HIM).

4.2.2 Phrase

Besides sentence stress we further perceive phonetic prominence along the waves (see figure 10).

There are essentially two types of waves: waves whose peaks reach or are close to the upper broken line and waves whose peaks are far away from it. The former we call 'major waves' and the latter 'minor waves'. When we examine the domains of these waves, each major wave usually contains one lexical morpheme and a few optional grammatical morphemes, while a minor wave mostly consists of only grammatical morphemes. In other words, the number of major waves corresponds roughly to the number of lexical morphemes contained in a given utterance with the following exceptions:

(a) A sequence of lexical morphemes such as a noun and an adjective:

moni fiitii *Fijian money*

In slow speech fiitii forms a major wave or two while moni forms a minor wave. In such a case a major wave seems to occur on the semantically more important morpheme in that particular utterance. In fast speech they form together one major wave.

(b) Monosyllabic personal pronouns such as ia *he* and au *I*. In fast speech they are often attached to the preceding or the following lexical morpheme to form a major wave. For example kave *take* and ia *he* in figure 10 form together a major wave.

The domain of a major wave is most closely related to the domain of a morphological phrase, so we call it 'phonological phrase', and the stress of this domain 'phrase stress'. A similar notion is called 'prosodic word', 'phonological word' or 'stress group' by Cutler and Ladd (1983) and 'Akzentgruppe' by Zinglé (1982).

A phrase stress occurs normally on the first syllable of a lexical morpheme, or on the final syllable, if it is the only long syllable in the morpheme.

//ka hInangaro// *will want*
//te pahII// *the ship*

However, before a phrase boundary which is followed by a pause, a phrase stress can occur on the final short vowel. In such cases the final short vowel is never devoiced, but slightly lengthened.

//te tangatA// *the man*
//te tamaiti inA// *the child there*

4.2.3 Measure

Besides the stress along the major wave, we also perceive stress along the minor wave, particularly in carefully spoken slow speech.

The domain of each minor wave corresponds most closely to the notion 'measure' introduced by Scott (1948) for Fijian 'words' and expanded by Schütz as a phonological unit for Fijian (1976), Hawaiian (1978, 1981) and Maori (1984).

The minimum unit of the domain of this minor wave in Penrhyn is one long syllable. The longest wave includes a lexical morpheme. While one lexical morpheme sometimes extends over two waves, a few grammatical morphemes can also form one wave (figure 11). Each wave has one stressed syllable. In other words, this is the minimal stress domain, and we also call it 'measure'. It is a phonological notion, which is placed between phonological phrases and syllables.

¦rAA ¦	UA ¦		*they* (dual)
¦p00 ¦	p0ngi ¦		*morning*
¦kI te ¦			*to the*
¦ei ¦	kIte ¦	m00na ¦	*to appear for him*

In careful slow speech, like in a one-word utterance, a measure stress occurs on the penultimate syllable or on the final long syllable.

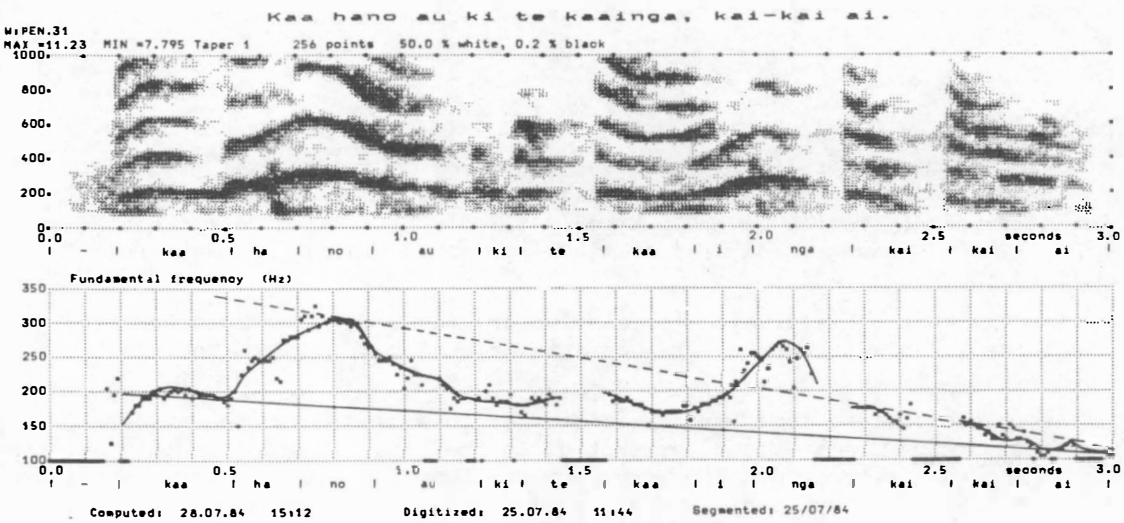


Figure 11: A Penrhyn declarative sentence, divided into measures

(Each measure corresponds to one wave.)

¦kaa ¦ hano ¦ au ¦ ki te ¦ kaa ¦ inga ¦ kai ¦ kai ¦ ai ¦

#kaa hAno au ki te kaaInga KAI-kai ai#
go I home eat
I am going HOME to have a meal there.

(Phrase stress is marked by capital letters. A sentence stress occurs on kaaInga *home*, where the fundamental frequency exceeds the upper broken line.)

We have not found any successful rules yet as to when a grammatical morpheme is attached to a preceding or a following lexical morpheme or when two or three grammatical morphemes can be combined to form a measure.

4.3 Interaction of stress in different levels

It is not yet clear how measure stress, phrase stress and sentence stress influence each other in an utterance. It seems that in carefully spoken slow sentences one perceives a measure stress throughout the utterance.

In normal fast speech, however, some measures are attached to a preceding or a following measure and a regrouping occurs. The new group is likely to act like a phrase and accordingly gets a phrase stress. Therefore a potential measure is suppressed (see kave ia and i aa au in figure 10). Both, phrase stress and

measure stress, are possible candidates for receiving sentence stress, depending upon how important the corresponding morphemes are in a given sentence.

This is why 'words' based on a morphological analysis sometimes receive stress on the penultimate syllable and sometimes on the first syllable. It is now very clear that there is no 'intrinsic word stress' in Penrhyn. A stress in a word (= lexical morpheme) can be shifted from one syllable to another according to its position within a measure, a phrase or a sentence, and depending upon the style or the tempo of the speech.

4.4 Function of stress

One perceives a certain rhythm along the waves of fundamental frequency, which appears to be a physiological necessity for language production as well as for perception. People become extremely tired of listening to synthetic speech, if suprasegmental features are lacking. This indicates that stress simplifies the understanding of speech.

It seems that a hierarchical tree structure based on the stress pattern can be formed. This tree structure consists of measures, phrases and sentences in phonological terms.

The tree structure is in principle similar to a syntactic tree structure of for instance a Chomsky-model, but not quite the same. Suprasegmental features are superimposed in the syntactic structure and function when they are necessary. For instance, where a syntactic ambiguity occurs, a phonological phrase boundary helps a speaker and a hearer decide on the right choice.

5. NOTATIONS

Illustrative texts are cited in phonemic transcription, modified in the following respects:

#	: sentence boundary
//	: phrase boundary
¦	: measure boundary
/ /	: phonemic transcription
[ ]	: phonetic transcription
upper case	: stressed syllable
hyphen	: morphologically determined complex word

NOTES

1. See section 5 for details of the notation.
2. Penrhyn has 10 consonant phonemes: p t k m n ng r v s h, and five vowel phonemes: i e a o u.

REFERENCES

BIGGS, B.G.

1961 The structure of New Zealand Maori. *Anthropological Linguistics* 3:1-54.

CUTLER, A. and D.R. LADD, eds

1983 *Prosody models and measurements*. Berlin: Springer.

DELMONTE, R.

1982 An automatic text-to-speech prosodic translator for the synthesis of Italian. *Fortschritte der Akustik-DAGA* '82:1021-1026.

FOLKINS, J.W., C.J. MILLER and F.D. MINIFIE

1975 Rhythm and syllable timing in phrase level stress patterning. *Journal of Speech and Hearing Research* 18:739-753.

HIROSE, K. and H. FUJISAKI

1983 Analysis and synthesis of voice fundamental frequency contours of complex sentences. *Proceedings of the 11th International Congress of Acoustics* 4.

HOHEPA, P.

1967 A profile-generative grammar of Maori. Ph.D. dissertation, Indiana University, Bloomington.

KOENIG, W., H.K. DUNN and L.Y. LACY

1946 The sound spectrograph. *The Journal of the Acoustical Society of America* 17:19-49.

LEHISTE, I.

1970 *Suprasegmentals*. Cambridge, Mass.: M.I.T. Press.

MARKEL, J.D. and A. GRAY

1976 *Linear prediction of speech*. Berlin: Springer.

NEWBRAND, H.L.

1951 A phonemic analysis in Hawaiian. M.A. thesis, University of Hawaii.

PAULUS, E. and E. ZWICKER

1972 Programme zur automatischen Bestimmung der Lautheit aus Terzpegeln oder Frequenzgruppenpegeln. *Acustica* 27:253-266.

SCHÜTZ, A.J.

1976 Fijian prosody I: syllables and groups. *Working Papers in Linguistics, University of Hawaii* 8/2:75-100.

1978 Accent in two Oceanic languages. *Anthropological Linguistics* 20/4: 141-149.

1981 A reanalysis of the Hawaiian vowel system. *Oceanic Linguistics* 20/1: 1-44.

1984 Accent and accent units in Maori: the evidence from English borrowings. *Journal of the Polynesian Society* (in press).

SCOTT, N.C.

1948 A study in the phonetics of Fijian. *Bulletin of the School of Oriental and African Studies* 12:737-752.

VAISSIÈRE, J.

1983 Language-independent prosodic features. In A. Cutler and D.R. Ladd, eds *Prosody models and measurements*, 53-66. Berlin: Springer.

WOLF, H.E.

1982 Intonationsmotive zur Prosodiesteuerung nach Regeln für die Sprachsynthese. *Fortschritte der Akustik-DAGA* '82:1031-1034.

YASUDA, A.

1968 The structure of the Penrhyn phrase. M.A. thesis, University of Hawaii.

ZINGLÉ, H.

1982 Automatische Prosodiesteuerung in der Synthese der deutschen Sprache. *Fortschritte der Akustik-DAGA* '82:1027-1030.

ZWICKER, E.

1982 *Psychoakustik*. Heidelberg: Springer.

ZWICKER, E. and R. FELDTKELLER

1967 *Das Ohr als Nachrichtenempfänger*. Stuttgart: S. Hirzel.

Yasuda-Graefe, A. and Graefe, V. "A computational approach to stress patterns in Penrhyn". In Geraghty, P., Carrington, L. and Wurm, S.A. editors, *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics.* **C-93:357-376**. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.357*

RESOURCES FOR AUSTRONESIAN LINGUISTIC RESEARCH IN THE NATIONAL LIBRARY OF AUSTRALIA, CANBERRA

Geraldine Triffitt

The strength of the National Library of Australia lies in its materials relating to Australia and its surrounding region, particularly Papua New Guinea, Indonesia, New Zealand and the Pacific. This paper will discuss some of the resources of interest to Austronesian linguists, particularly early translations into vernacular languages and 19th century linguistic studies in these areas.

I have worked at the Library for nearly 12 years during which time I have been a heavy user of the collections both as a bibliographer and as a linguistics student. Although all readers can access most of the material through the various catalogues, I have the advantage of being able to browse among the stacks and through serendipity find items of interest.

After briefly describing the history of the Library I shall mention some collections of interest to Western Austronesian linguists before I concentrate on Pacific language materials.

HISTORY OF THE NATIONAL LIBRARY OF AUSTRALIA

After Federation in 1901, the Commonwealth Parliament was still located in Melbourne, Canberra having not yet been conceived let alone born. After receiving advice from the state librarians, the Joint Parliamentary Library Committee opted for the development of a truly national collection on the lines of the Library of Congress, Washington. Their report of 1903[1] stressed the importance of securing and preserving all works and documents connected with the discovery, settlement and early history of the various States of the Commonwealth. By its nature such material would have a considerable Pacific component.

The Australian collection was boosted as a result of certain legislation. In 1911 by Act of Parliament[2] the Library accepted the Petherick Collection carefully assembled by Edward Petherick and rich in Australiana and Pacificana. The following year the Copyright Act[3] required Australian publishers to deposit a copy of each publication in the Parliamentary Library, thereby ensuring the maintenance of a national collection.

In the 1920s the Commonwealth Parliament moved from Melbourne to Canberra. The non-Parliamentary collection of the Library was renamed the Commonwealth National Library and separately housed. Finally in 1960 with the passing of the

---

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics,* 377-403. *Pacific Linguistics*, C-93, 1986.

National Library Act,[4] the two libraries were separated. Under the Act, the new National Library of Australia was empowered to maintain and develop a national collection of library material, including a comprehensive collection of library material relating to Australia and the Australian people.

When the National Library moved out of its warehouses, basements and other temporary storage into the new Greek-inspired building in the Parliamentary triangle in 1968, the collection totalled more than a million books and thousands of non-book materials including manuscripts, maps, aerial photographs, pictures and moving picture films. This has since risen to four million volumes including microforms.

INDONESIAN COLLECTION

The Second World War showed up Australia's ignorance of Asia especially South-East Asia. The resources for strategic planning were very meagre. In the postwar years as a measure of rectifying this the National Library undertook a vigorous program of acquiring formed collections particularly of Indonesian materials in Indonesian and Dutch. For example in 1959, the Library purchased the collection of R.A. Kern, son of J.H.C. Kern after whom the Kern Institute in Leiden, Netherlands, was named. The 2000 items, primarily in Dutch, related to Indonesia in the 19th century. Associated with the collection were transcripts of Sundanese manuscripts, particularly texts of Sundanese folk songs. These were copied for Mr Kern, who lectured in the Sundanese language at Leiden University, from the Sundanese manuscript collection of the Bataviaasch Genootschap voor Kunsten en Wetenschappen. However as some of the original manuscripts are believed to have been lost, the transcriptions may be the only extant copies.

Two collections are of interest to Batak scholars. The library of Professor Tobing, a prominent Indonesian scholar in the fields of theology and anthropology, was purchased in 1965. Included among the linguistic material was the only known existing dictionary of the Simalungun dialect compiled by J. Wisner Saragih in 1936 (Brakel 1978:16). In 1978 the National Library acquired photocopies of transliterations of Batak texts written on *pustahas* (tree-bark books) in the Museum Pusat in Jakarta. Some of these texts on magic, divination and medicine were written in the Mandailing dialect probably before 1850 since the Mandailing people were converted to Islam in the first half of the 19th century (Voorhoeve 1978). The Korn collection on microfiche included printed and manuscript material collected by Professor Korn, an authority on adat law. About half the collection which was filmed at the Royal Institute of Anthropology, Leiden, was devoted to Bali, much of it transcripts of Balinese texts. Other documents were in different languages particularly Minangkabau, Nias, Batak and Djambi.

The Library has three 17th century Malay language books. One by David Haex, published in 1631, is a Malay-Latin and Latin-Malay dictionary which includes Amboyna, Banda and Moluccan words and phrases. A translation of Genesis published in The Hague in 1662 has parallel text in Dutch and Indonesian (Malay). The third book is a translation of the four Gospels and Acts of the Apostles by Thomas Hyde published in Oxford in 1677.

On a more contemporary note, the New York Public Library donated 6,000 Indonesian monographs published since Independence. These included texts in regional languages particularly Minangkabau, Achinese, Balinese and Batak, and studies by H.G. Tarigan on Karo and Simalungun dialects.

Since 1971 the Library has had an Acquisition Office in Djakarta staffed by an Indonesian-speaking staff member. This office collects a basic set of Indonesian publications for several other Australian libraries under a cooperative scheme. The National Library collects Indonesian material comprehensively and lists the material in the *Indonesian Acquisition List*. As a result of all this collecting activity the National Library of Australia has one of the largest collections of Indonesian materials in the world.

OTHER WESTERN AUSTRONESIAN MATERIALS

Western Austronesian languages are represented in other formed collections. The working library of Professor H. Otley Beyer, Foundation Professor of Anthropology at the University of the Philippines, was acquired in 1972. Included in his pamphlet collection are short literary works in vernacular languages of the Philippines, phonemic studies of local languages and articles and journal reprints of language studies.

In 1959 the Library purchased a collection of over 600 linguistic items among which were the copies of Sundanese manuscripts and Kern's lecture notes on the Sundanese language. The collection comprised basic works in most Austronesian languages and dialects from Madagascar to Tahiti. The main emphasis was on the languages of Indonesia and the Philippines with a few items from Oceanic languages. Altogether the collection contained 58 grammars and 91 dictionaries largely published in the late 19th and early 20th centuries. There are some postwar works, for example, Milner's *Fijian grammar* published in 1956.

It is thought that the collection was originally Kern's; the Dutch bookseller did not name the owner but only noted that he had spent many years in the East. The Indonesian-Filipino thrust of the collection is exemplified by Gajo-Dutch and Favor-Dutch dictionaries, Javanese grammars, Hiligaina dialect and Bikal grammars and Isneg texts to mention just a few. There are several studies of the Malagasy language including Dahl's *Malgache et maanjan: une comparaison linguistique*, published in Oslo, 1951. The collection has been fully catalogued and has been integrated into the Library's general linguistic stock.

PACIFIC LANGUAGE MATERIALS

I will now turn to the main thrust of the paper: the Pacific language materials. Most of the linguistic studies and vernacular texts are associated with the Library's three great collections of Australiana and Pacificana: the Petherick, Ferguson and Nan Kivell collections.

PETHERICK COLLECTION

Edward Augustus Petherick, who worked in the book trade in London for Francis Edwards and also for George Robertson of Melbourne, collected items to be recorded in his "Bibliography of Australia and the Pacific". This was never published but remains as loose sheets in pamphlet boxes in the Library's Petherick Reading Room. Out of the 92 boxes, 26 contain material relating to the Pacific. For linguists, boxes 26 and 27 contain entries for Pacific languages written by hand on slips of paper pasted onto quarto sheets. Some of the entries are references to works

mentioned in newspapers and contemporary publications, others are analytical entries from works that he possessed. Of particular interest are the many references to the linguistic notes of explorers in the 18th and early 19th centuries. Petherick had a comprehensive collection of accounts of voyages of explorers in the Pacific. Many of these, including Arago, Cook, Dumont D'Urville, D'Entrecasteaux, Labillardiere, Lesson, Moerenhout and the United States Exploring Expedition recorded the speech of the inhabitants of the islands they visited. Some expeditions included linguists, such as Gaimard on the *Astrolabe* with Dumont D'Urville, and Horatio Hale with the United States Exploring Expedition.

Petherick offered his collection of 10,000 volumes and 6,500 pamphlets, maps, manuscripts and pictures to several institutions, but it was finally accepted by the Commonwealth of Australia in 1909. It provided the nucleus of the stock of the Parliamentary Library listed in *Catalogue of the books, pamphlets, pictures and maps in the Library of Parliament to September, 1911*. Among the linguistic material are early dictionaries and grammars of Maori, Tongan, Hawaiian, Gilbertese, Fijian, Samoan, Tahitian and other languages of the French colonies in the Pacific. Some items are very rare such as the Rev. Nathaniel Turner's *First lessons in the languages of Tongataboo* published in 1828 which was probably the first book used in Tonga. Books in the Petherick Collection have been integrated into the general stock of the National Library. A separate collection is the Petherick pamphlets housed in filing cabinets under broad subject headings. Included in the items filed under Philology are reprints of papers by Sidney Ray on Melanesian languages. The catalogue of the Library of Sir George Grey relating to Philology provides a bibliography of early works in Melanesian, Polynesian, Papuan and Australian languages.

REX NAN KIVELL COLLECTION

This collection, first transferred to the Library in 1959, has combined pictorial, manuscript and printed material for a record of European man in the Pacific, his discoveries, exploration and settlement. Although it is particularly strong in pictorial material, there are some valuable early Pacific language items particularly in Maori. There are at least 50 items mentioned in Williams *A bibliography of printed Maori to 1900*. Many are Bible translations but other items in the Nan Kivell collection include government proclamations in Hawaiian and Maori, programs of Tahitian canoe racing and the laws of England and adventures of Robinson Crusoe in Maori. Some historical personages are represented, notably Sir George Grey, Governor of New Zealand, and Kamehameha III, King of the Hawaiian Islands. Their letters, proclamations and a work by Grey on the mythology and traditions of the New Zealanders in Maori are included. Nan Kivell was born in Christchurch, New Zealand, and while he collected art for the Redfern Gallery in London he pursued his hobby of collecting material relating to the South Seas. He combined both interests in his ambition to compile a dictionary of portraits of people connected with the Pacific, the originals for which are also in the Library. The Nan Kivell books have been kept as a special collection, supplementing the pre-1850 imprints collected by Sir John Ferguson and housed in the Ferguson Room for rare Australiana and Pacificana.

FERGUSON COLLECTION

With the purchase of the remainder of the collection of Sir John Ferguson's library after his death in 1969, the Library acquired its largest private

collection, numbering 34,000 items. Sir John concentrated on book collecting in specific fields including the languages of the Pacific and the history of missions in the area. As the son of a Presbyterian minister and legal advisor to the Presbyterian Church he was able to establish contact with missionaries in the Pacific, particularly New Guinea and the New Hebrides, and to assemble a wealth of translations of the scriptures and other vernacular linguistic material, many representing the first printing in the area.

The Island Language Collection comprises some 2,300 titles mostly originating from missions. The pamphlets, many of which are in a fragile condition, are filed in envelopes by Dewey Classification in filing cabinets. This material was published after 1850 and includes some comparatively recent vernacular items. The majority of these pamphlets are in Melanesian languages with a strong bias towards the languages of Vanuatu.

Most of the books are in compactus shelving in the basement. It is rather exciting to pull back the compactus shelves and be confronted with 34 shelves of 19th century Pacific dictionaries and Bible translations. It is a thrill to handle a volume previously owned by a famous missionary, inscribed with his signature and annotated in his own handwriting. One such find was a copy of the second edition of Hazlewood's Fijian and English dictionary and grammar with the signatures J.H.L. Waterhouse on the endpapers and Rev. J.P. Chapman, Fiji on the title page. The dictionary was annotated with additional entries, but the endpapers had a wealth of information of interest to linguists and anthropologists - a list of old Bauan numerals, references to Pritchard's reminiscences, vessels in Fijian waters December 1844 to January 1850 and Tahitian ceremonial language. Of more local interest were Cakaudrove words and their meanings, a description of the Dakuwaqa legend and alternative names for the Shark God, items of interest about Somosomo, Taveuni, botanical references including a description of *Medinilla waterhousei* found on the mountains above Somosomo with a note in another hand "Is it Tagimaucia?" (a flower unique to this area). Other interesting items I found were a Tahitian translation of the Gospel according to St Matthew, printed by the Windward Mission Press in 1820 and believed by Ferguson to be the earliest book with a Tahitian imprint, translations of the Acts (1822) and St Mark's Gospel (1827) in Tahitian and the Rev. John Davies' Tahitian grammar of 1823. There was *Ka Palapala hemolele* the complete Bible translated into Hawaiian in 1843 and Lorin Andrews *Dictionary of the Hawaiian language*, 1865 with a bookplate Bibliotheca Lindesiana. This collection is presently being catalogued for entry into the National Bibliographic Database and in some cases material is being relocated to more appropriate locations. Notes by Ferguson himself and by Pauline Fanning, formerly responsible for the Australian and Pacific collections at the National Library, give the provenance of some of the volumes. One note found in Hazlewood's *Feejeean-English dictionary* of 1850 stated it had been bound by a beachcombing bookbinder in Fiji, who stayed at the mission, went home to get married, and returned to Fiji presumably to lead a righteous and sober life there. Among the Maori materials collected by Ferguson are a set of 11 pamphlet boxes containing over 200 items with pre-1900 imprints arranged by Williams number.

Ferguson manuscripts

In addition to the published works collected by Sir John Ferguson, which formed the basis of his *Bibliography of Australia*, his collection of manuscripts, news cuttings, correspondence and miscellanea are kept in 60 boxes in the

Manuscript Section. There is a brief listing of the contents of these boxes, and although several have items of Pacific interest, box 16 is particularly interesting to linguists. It contains correspondence between Ferguson and missionaries requesting copies of books in vernacular languages. These request letters and the replies from the missionaries give interesting personal insights into the lives of mission families on isolated islands of the Pacific. In addition, box 16 contains two bibliographies compiled by Ferguson of works about the New Hebrides, one a "bibliography of New Hebrides language books", which is arranged by the location of the language, the other a manuscript "Bibliography of New Hebrides Islands" on single sheets arranged chronologically in three parts. The latter is not language oriented, but is valuable for studies of missions. Other linguistic bibliographies are located among the bibliographies in box 24. There are two typescript bibliographies compiled by Howard Malcolm Ballou in 1908: "Bibliography of books in the native Hawaiian language" and "Preliminary bibliography of books in the languages of the Micronesian islands evangelized by missionaries of the Hawaiian Board of Missions". Another bibliography "Works in native languages translated by missionaries of the Australasian Wesleyan Methodist Missionary Society" is divided into nine parts - two for New Britain, five for Papua, one each for New Ireland and the Solomon Islands.

OTHER MANUSCRIPTS

The Library has very strong resources for the studies of missions in the Pacific and Australia. Some of the records of missionary societies and individual missionaries have been deposited in the Manuscript Section, and where originals are held elsewhere microfilm copies of records are available. Among original manuscripts are records of the Methodist Missionary Society, Wesleyan Methodist Mission, Tonga and the papers of J.G. Paton (New Hebrides), James Buller (New Zealand), letters of Richard Fallowes (Solomon Islands), Daniel MacDonald (New Hebrides), Platt Family (Raietea, Society Islands) and John Thomas (Tonga). The Library purchased some early letters written in Tahitian by Taero from Maupiti in 1827 and Mai, a chief of Bora Bora, 1825, to the London Missionary Society, London. Other manuscript collections contain linguistic material. Gerald Maxwell, 1898-1959, formerly Chairman of the Native Lands Commission in Fiji was interested in Fijian mekes and songs, copies of which I found among his manuscript collection. He also collected some Fijian genealogical tables. J.H.L. Waterhouse's vocabulary notebook of Solomon Island words was presented to the National Library in 1943 by his widow.

MICROFILMED MATERIAL

The National Library of Australia has been deeply involved in several major projects for microfilming Pacific material. The Pacific Manuscripts Bureau and the Australian Joint Copying Project make available to researchers an abundance of Pacific material on microfilm much of it of interest to linguists. As these microfilms are widely available I will only briefly mention some of the linguistic material.

The Pacific Manuscripts Bureau is a cooperative scheme between four libraries with Pacific research collections: the Mitchell Library, National Libraries of Australia and New Zealand and the University of Hawaii Library. The project

office is at the Research School of Pacific Studies at the Australian National University. As a result of filming mission records and diaries of explorers many wordlists, grammars and vernacular newspapers have become available. Some examples are Crawford's Tuvaluan material including wordlists and lessons in the language (PMB 919), a 23-page Fijian vocabulary notebook written by Captain W.P. Richardson in 1810-1812, copies of the newspapers *Taumua lelei* 1929-1939 from Tonga, *Te vea Maohi*, a Tahitian monthly, 1936-1944 and *Ai Tukutuku vakalotu* from 1897-1935, a Fijian journal published by the Methodist Mission. (There are originals of *Ai Tukutuku vakalotu* in the Ferguson collection housed in pamphlet boxes in the basement.) The monthly journal *Pambu* provides information about newly filmed documents and indexes to previous reels.

In 1948 the National Library of Australia and the Mitchell Library started filming British government records pertaining to Australia and the United Kingdom as part of the Australian Joint Copying Project. Later Phyllis Mander-Jones identified other manuscripts in Britain relating to Australia, New Zealand and the Pacific which have since been microfilmed and indexed by the National Library for publication in the *A.J.C.P. Handbooks* (Mander-Jones 1972).

Mission records which have been microfilmed include those of the London Missionary Society and Methodist Missionary Society. One particular linguistic collection that was filmed was Sidney Ray's collection of manuscripts in the Library of the School of Oriental and African Studies, London, including some of his Melanesian vocabularies. A guide to the collection is "Handlist of manuscripts in the Sidney Herbert Ray Collection". Some Pacific vocabularies are included such as Guadalcanal dialect, Mortlock Islands, and Chamorro on Guam. Fijian specialists would be interested in Lelean's wordlist from different regions in Fiji, and the vocabularies of Somosomo, Macuata, Kadavu, Bua, Nadroga and Ba to mention just a few.

OCEANIC LINGUISTIC PAMPHLETS

In 1979 the Library purchased a collection of 200 pamphlets, reprints and journal articles on linguistics and anthropology, formerly in the possession of Sidney Ray. The geographical extent of the subjects range from Africa to Canada, but the majority of items are concerned with Oceania in the late 19th and early 20th centuries. Although much of the material is freely available it is nevertheless an interesting assemblage of items drawn from sources much wider than linguistic journals and including scientific and geographical journals. These pamphlets are bound in 13 volumes and shelved in the main linguistic stacks.

ACCESS TO MATERIAL

How does one obtain this material? The National Library is a closed-access library, so access is through the catalogues. The Dictionary catalogue is a single sequence catalogue which contains entries for the library stock acquired up to 1966. Most of the island language material may be found there. After 1966 the four-part divided catalogue consisting of the Name file, Subject file, Serials and In Process files prevailed until March 1980. Since then the catalogue has been automated and a fiche catalogue is available to readers. Cataloguing projects are underway to enter the Nan Kivell, Petherick pamphlets and Ferguson collections into the computer, so that records are incorporated into the National Bibliographic Database. With the exception of some special collections kept

separate, such as the Nan Kivell and the Ferguson, materials in the general collections are arranged according to the Dewey Decimal Classification. Some Ferguson materials are arranged in the order given in his *Bibliography of Australia*, and others are in separate Dewey sequences. The Manuscript Section has its own catalogue and material is available to scholars in accordance with the access provisions of the donors.

The Petherick Reading Room on the first floor is the service centre for the Australian and Pacific collections. This is the location of the Petherick bibliography and has a number of published bibliographic tools. Materials in South-East Asian languages particularly Indonesian are available from the Area Studies Reading Room on the Lower Ground 1 floor. This area is also the closest point to the general linguistic materials stack. The Newspaper Reading Room on the same floor services both newspapers and microforms, and has guides to the PMB and AJCP microfilms as well as a separate catalogue of microform items. Unfortunately much of the material I have mentioned is not available for inter-library loan, but items can be copied.

PRESENT DEVELOPMENTS

In the past the Library has benefitted from the endeavours of some inspired private collectors. What of the future? *The National Library of Australia selection policy* published in 1981 states:

> The Library gives high priority to the collection of Pacificana. It collects comprehensively publications originating in or relating to the South Pacific (including New Zealand, Antarctica and Papua New Guinea).[5]

Pacificana is given the second highest priority after Australiana. Similarly in respect to selection of material by language, the policy is stated:

> Publications in the languages of New Zealand, the islands of Micronesia, Polynesia, Melanesia and Papua New Guinea are collected in depth.[6]

In 1981 the Committee on Pacific Region Acquisitions (COPRA) was established as a cooperative effort between the Australian National University Library and the Research School of Pacific Studies of that institution and the National Library of Australia. Surveys of the resources at both institutions were undertaken as a preliminary to the rationalisation of Pacific material. In particular attempts are being made to establish contacts and to maintain exchange arrangements already existing with a view to increasing the supply of government publications from Pacific countries.

BIBLIOGRAPHICAL GUIDES

Although most of the Pacific language material is catalogued or is in the process of being catalogued, it may be located in a formed collection or in the general linguistics stacks. The Library has published several guides and finding lists which are useful for this reason. *The Guide to the collections,* compiled by C.A. Burmester, is an indispensable four-volume treasurehouse of information about the collections in the Library. It gives a description of a large range of subjects and their location and access. Mr Burmester played an important part

in the acquisition of much of the National Library's stock and his essay "History of the National Library's collections" in *National Library of Australia selection policy* makes fascinating reading. He also summarised the relevant sections of his guide in *Christian missions and missionaries in the Pacific: brief notes on research materials in the National Library of Australia, 1977*. A finding list for newspapers in Pacific vernacular languages is *New Zealand and Pacific Islands newspapers held in the National Library of Australia*, 1981. Pauline Fanning's article on the Ferguson collection in *Australian Library Journal*, 1969, gives a detailed account of the island language material and missionary records in the Ferguson collection.

The accompanying bibliography of Austronesian languages published before 1840 in the National Library of Australia lists some of the resources for individual language studies.

ACKNOWLEDGEMENTS

This was a private after-hours project and I am indebted to the help given me by my colleagues at the National Library of Australia. In particular, I wish to thank Pauline Haldane and Ilse Soegito from Area Studies, Lesley Bezear for permission to copy the shelf list, Elisabeth Jovanovic for keeping me informed on the Nan Kivell collection, Margaret Brennan and the Petherick Reading Room staff and the staff of Manuscripts Section, Rare Books and Newspapers. Pauline Fanning's encyclopaedic knowledge of the collections guided me to track down elusive items and thereby often opened up new avenues to explore. Finally, I dedicate this paper to my late husband, Tom, who edited it, and to whom I owe my love of the Pacific and my interest in bibliography.

NOTES

1. *Report from the Library Committee (sitting in conference with the Library Committee of the Senate)*. Melbourne: Robert S. Brain for the Government of Australia, 1903 (Australia. Parliament. House of Representatives, Paper no.5, 1903).
2. Petherick Collection Act, 1911 (No.4, 1911).
3. Copyright Act 1912 (No.20, 1912).
4. National Library Act 1960 (No.69, 1960).
5. 1981:5.
6. 1981:105.

REFERENCES

AUSTRALIA. Parliament. House of Representatives

1903 *Report from the Library Committee (sitting in conference with the Library Committee of the Senate)*. Robert S. Brain, for Government of the Commonwealth of Australia.

BRAKEL, Lode

1978 Batak texts in the National Library. Southeast Asian Research Materials Group *Newsletter* 11:15-17.

BURMESTER, C.A.

1977 *Christian missions and missionaries in the Pacific: brief notes on research materials in the National Library of Australia*. Canberra.

BURMESTER, C.A., comp.

1974-1982 *National Library of Australia: guide to the collections*. 4 vols. Canberra: National Library of Australia.

FANNING, Pauline

1962 The Australasian collection of Mr Rex Nan Kivell, in the National Library of Australia, Canberra. *Australian Library Journal* 11:139-142.

1969 The Ferguson collection. *Australian Library Journal* 18:341-349.

LANGDON, Robert

1969 Bibliography of Australasia by E.A. Petherick. *Journal of Pacific History* 4:154-158.

1970 Pacific Manuscripts Bureau. *Journal of Pacific History* 5:159-163.

MANDER-JONES, Phyllis, ed.

1972 *Manuscripts in the British Isles relating to Australia, New Zealand, and the Pacific*. Canberra: Australian National University Press.

NATIONAL LIBRARY OF AUSTRALIA

1981a *National Library of Australia selection policy*. Canberra: National Library of Australia.

1981b *New Zealand and Pacific Islands newspapers held in the National Library of Australia*. Canberra: National Library of Australia.

VOORHOEVE, Petrus

1978 Tree bark books of Sumatra. Southeast Asian Research Materials Group *Newsletter* 10:11-12.

BIBLIOGRAPHY OF WORKS ON AUSTRONESIAN LANGUAGES PUBLISHED BEFORE 1840 IN THE NATIONAL LIBRARY OF AUSTRALIA, compiled by GERALDINE TRIFFITT

The primary aim of this bibliography was to give examples of the material available to Austronesian linguists in the Ferguson, Nan Kivell and Petherick collections. This was supplemented by items from the catalogue, particularly Indonesian and Malay items.

The sources of the citations were the shelf-list cards for the Oceanic language monographs which I copied while undertaking research for my conference paper, and the unpublished Bibliography of Australasia compiled by Petherick mostly from his own collection.

The majority of the items cited are either wordlists compiled by exploring expeditions, or translations of the Bible into vernacular languages, and not general linguistic studies - which are a resource that I have yet to research.

Not all items have been sighted. Where items have not been seen, bibliographic descriptions have been checked in the following sources:

AUSTRALIA. Parliament. Library

1912 *Catalogue of the books, pamphlets, pictures, and maps in the Library of Parliament to September 1911.* Melbourne: Government Printer.

FERGUSON, *Sir* John Alexander

1975-1977 *Bibliography of Australia.* Canberra: National Library of Australia. 7 vols. Facsimile reprint.

KLIENEBERGER, H.R., comp.

1957 *Bibliography of Oceanic linguistics.* London: Oxford University Press.

NATIONAL LIBRARY OF AUSTRALIA

1977 *Select bibliography of voyages and travels, 1730-1830.* Canberra.

O'REILLY, Patrick and Edouard REITMAN

1967 *Bibliographie de Tahiti et de la Polynésie française.* Paris: Musée de l'Homme.

PETHERICK, Edward Augustus

1500-1917 Bibliography of Australasia. MS boxes of loose-leaf material. Canberra: National Library of Australia.

SHARP, Andrew

1960 *The discovery of the Pacific Islands.* Oxford: Clarendon Press.

SNOW, Philip A.

1969 *A bibliography of Fiji, Tonga and Rotuma: preliminary working edition.* Canberra: Australian National University Press.

WILLIAMS, Herbert William

1924 *A bibliography of printed Maori to 1900.* Wellington: Government Printer.

GENERAL

HERRERA Y Tordesillas, Antonio de

1622 Dictionaire du langage des isles de Salomon; de l'isle de Cocos, de Nova Guinea, de l'isle Moyse, de l'isle de Mao. In his *Description des Indes Occidentales, qu'on appelle aujourdhuy le Nouveau Monde*, 196-199. Amsterdam.

RBq JAF 10

DALRYMPLE, Alexander

1771 Vocabulary of languages in some of the islands visited by Le Maire and Schouten. In his *Historical collection of the several voyages and discoveries in the South Pacific Ocean*, vol.II, 7p. London.

N+ 910.4 D151

COOK, James

1773 [Comparison of languages of Prince's Island, Java and Malacca, South Sea and Madagascar.] In John Hawkesworth *An account of the voyages for making discoveries in the southern hemisphere ...*, vol. III, 371-374. London.

RBNq COO

ANONYMOUS

1774 Vocabulaires: tirés des langues barbares de divers Peuples Austraux. In M. Richer *Histoire moderne ...*, vol.25, 419-436. Paris: Saillant and Nyon; Veuve Desaint.

(Langues 'des Isles Salomon, de l'isle Moyse, de l'isle Moa, près des côtes de la Nouvelle Bretagne'.)

RB 909 MAR

COOK, James

1777 A table exhibiting at one view specimens of different languages spoken in the South Sea from Easter Island westward to New Caledonia. In his *Voyage towards the South Pole and round the world*, vol.II, between 364 and 365. London.

RBN 910.4

FORSTER, John Reinhold

1778 Comparative table of 46 words in the various languages in the Isles of the South-Sea and of various nations to the East and West of it. In his *Observations made during a voyage round the world on physical geography, natural history and ethic philosophy ...*, 284. London.

RBq 504

LE MAIRE, Jacob

1806 Comparison of Malagash, Lampoon County (district of Sumatra) and Cocos Island and discussion of language origins. In James Burney *A chronological history of the discoveries in the South Sea or Pacific Ocean*, vol.II, 445-447. London: Luke Hansard, Printer.

F372

WESTERN AUSTRONESIAN

INDONESIAN AND MALAY LANGUAGES

HAEX, David

1631 *Dictionarium Malaico-Latinum et Latino-Malaicum; cum aliis quamplurimis que quarta pagina edocebit.* Romae: typis et impensis Sac.

Congr. de Propag. Fide. (Includes Ambonese, Banda and Moluccan words and phrases.)
RB 499.203 HAE

BIBLE. O.T. Genesis. Indonesian

1662 *Genesis atau kitab deri miarahan kitab Mosis nang bermoula: bersalin dalam bassa Maleya deri pada pandita Daniel Brouwerius, dauloo pandita de negri Hellevoet-Sluys, de Negri Holland commediën de negri Banda daan Ambon de dalam India nang ca Timor.* 'S Gravenhage: by de erfgenamen van Wylen Hillebrandt Iacobsz van Wouw. (Parallel text in Dutch and Indonesian (Malay).)
RB 499.2 BIB

CRAWFURD, John

1823-1825 *De Indische archipel: in het bijzonder het eiland Java: beschouwd in de zeden, wetenschappen, talen, godsdienst, beschaving, koloniale belangen en Koophandel van derzelver inwoners: uit het Engelsch vertaald.* Haarlem: Loosjes.
919.1

ANONYMOUS

1833 *A VOCABULARY of the English, Bugis, and Malay languages, containing about 2000 words.* Singapore: Mission Press. (Compiled by G.H. Thompson?)
499.103 VOC P

Celebes languages

GAIMARD, M.

1834 Vocabulaire de la langue des Haifours de Manade (Celebes). In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie,* vol.II, 193-194. Paris: Ministère de la Marine.
910.4

Javanese

GERICKE, J.F.C.

1831 *Eerste gronden der Javaansche taal; benevens Javaansch leer- en lees boek, met eene woorden lijst ten gebruike bij hetzelve.* Batavia: Lands Drukkerij.
499.2205

Malay

BIBLE. N.T. Gospels. Malay

1677 *Jang ampat Evangelia ... daan Berboatan derri jang Apostoli bersacti, bersalin dallam bassa Malayo = That is the four Gospels ... and the Acts of the Holy Apostles, translated into the Malayan tongue.* Oxford: Hall, Printer.
RB 226 BIB

WERNDLY, George Henrik

1736 *Maleische spraakkunst, uit de eige schriften der Maleiers opgemaakt.* Amsterdam: Wetstein.
499.105 WER

LABILLARDIERE, Jacques Julien Houton de

1800 Vocabulaire: Malais. In his *Relation du voyage à la recherche ... Tome second,* 3-43. Paris: H.J. Jansen.
F307

ANONYMOUS

1810 *A comparative vocabulary of the Burma, Malâyu and T'Hâi languages.* Semampore: Mission Press. 495.8

MARSDEN, William

1812 *A dictionary of the Malayan language in two parts, Malayan and English and English and Malayan.* London: Cox, Printer. 499.103

1812 *A grammar of the Malayan language.* London: Cox, Printer. 499.1

WERNDLY, George Henrik

1823 *Maleische spraakkunst.* [*2de druk*]. Batavia: Lands Drukkerij. 499.105 WER.2

ROORDA VAN EYSINGA

1824 *Nederduitsch en Maleisch woordenboek.* Batavia: Lands Drukkerij. (Indonesian words in Latin and Arabic script.) 499.203 ROO

1825 *Maleisch en Nederduitsch woordenboek onder goedkeuring en begunstiging der Hooge Regering van Nederlandsch Indië verwaardigd en uitgegeven.* Batavia: Lands Drukkerij. 499.203 ROO

BURDER, George

1826? *Rewajat Melajuw: tersalin dan tertara deri pada rewajet*^{2} *dan bahasa 'awrang Engris.* Molukhos: 'awleh J Kam, pandita, di-Ambon. 499.2

BRUIJN, R. le

1828 *Kitab njanji*^{2}*-an: jang sudah detambahij dengan barang mazmur*^{2} *dan tahlil*^{2} *indjil ...* Batawijah: di-patara'an Karadja'an. 499.2

1834 *Pengadjaran akau perboeatan Allah ...* Batavia: tertara di Kampong Parapattan. 499.2

SCHLEIERMACHER, Andreas August Ernst

1835 Grammaire malaie. In his *De l'influence de l'écriture sur le langage*, 409-710. Darmstadt. 414

Moluccan languages

ARAGO, Jacques Etienne Victor

1822 Vocabulaire: naturels de Guébé. In his *Promenade autour du monde, pendant: les années 1817, 1818, 1819 et 1820 ...*, 455-462. Paris. N 910.4

GAIMARD, M.

1834 Vocabulaire de la langue des habitans de l'île Guébé (Iles Moluques). In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie*, vol.II, 157-160. Paris: Ministère de la Marine. 910.4

Waigeo Island language

LABILLARDIERE, Jacques Julien Houton de

1800 Vocabulaire du langage des naturels de Waygiou. In his *Relation du voyage à la recherche de La Pérouse ... Tome second*, 66-69. Paris: H.J. Jansen.
F 307

D'ENTRECASTEAUX, Joseph-Antoine Raymond Bruni

1808 Vocabulaire de la langue des habitans de l'île Waigiou. In his *Voyage d'Entrecasteaux, envoyé à la recherche de La Pérouse*, vol.I, 585-590. Paris: Imprimerie Impériale.
F 461

ARAGO, Jacques Etienne Victor

1822 Vocabulaire: Alifourous, ou indigènes de Waigiou. In his *Promenade autour du monde, pendant les années 1817, 1818, 1819 et 1820 ...*, 462-463. Paris.
N 910.4

GAIMARD, M.

1834 Vocabulaire de la langue des Papous de Waigiou (Terre des Papous). In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie*, vol.II, 152-156. Paris: Ministère de la Marine.
910.4

MALAGASY LANGUAGES

CHAPELIER, M.

1833 Introduction au vocabulaire Madekass. In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie*, vol.I, 1-48. Paris: Ministère de la Marine. (Includes grammar notes first published in 1827 by M. Lesson.)
910.4

GAIMARD, M.

1833 Dictionnaire de langues Français et Madekass; vocabulaire Madekass -Français. In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie*, vol.I, 49-363. Paris: Ministère de la Marine.
910.4

MICRONESIAN LANGUAGES

Chamorro

CHAMISSO, Adelbert von

1821 Vocabulary of the dialects of Chamori (Mariana Islands) and of Eap, Ulea and Radack. In Otto von Kotzebue *Voyage of discovery into the South Sea and Beering's Straits ... in 1815-1818 in the ship Rurick*, vol.II, 409-433. London.
910.4

ARAGO, Jacques Etienne Victor

1822 Vocabulaire: les Chamorres ou marianais. In his *Promenade autour du monde ... Tome second*, 471-475. Paris.
N 910.4

GAIMARD, M.

1834 Vocabulaire de la langue des habitans de Gouaham (Iles Mariannes). In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie*, vol.II, 190-192. Paris: Ministère de la Marine. 910.4

Kusaiean

LESSON, R.P.

1825 Notice sur l'Ile de Oualan ou Strong. In *Nouvelles annales des voyages de la géographie et de l'histoire*, 81-91. Paris: Libraire de Gide Fils. (270 words in French and Oualan.) N 919.66

GAIMARD, M.

1834 Vocabulaire Français-Ualan. In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie*, vol.II, 175-181. Paris: Ministère de la Marine. 910.4

Palauan and Caroline Islands languages

KEATE, George

1788 A vocabulary of Pelew language. In his *An account of the Pelew Islands situated in the Western part of the Pacific*, 365-378. London. RBq 919.66 KEA

HOCKIN, John Pearce

1803 Vocabulary of the Pelew language. In George Keate *An account of the Pelew Islands*, 5th edn, supplement, 65-72. London. 919.66 KEA

ARAGO, Jacques Etienne Victor

1822 Vocabulaire: Iles Carolines. In his *Promenade autour du monde ...*, 476-492. Paris. N 910.4

GAIMARD, M.

1834 Vocabulaire de la langue des habitans de l'île Satawal (Iles Carolines). In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie,* vol.II, 182-189. Paris: Ministère de la Marine. 910.4

LÜTKE, Fedor

1835 ... de la langue des habitants de l'archipel des Carolines avec vocabulaire comparatif de quelques dialectes Carolinois. In his *Voyage autour de monde*, vol.II, 347-380. Paris. N 910.453 L776

O'CONNELL, James F.

1836 Glossary of the language of the Carolines. In his *A residence of eleven years in New Holland and the Caroline Islands ...*, 261-265. Boston: B.B. Mussey. (Vocabulary of Ponape.) F 2163

MELANESIAN LANGUAGES

Fijian

GAIMARD, M.

1834 Vocabulaire Français-Viti. In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie*, vol.II, 137-142. Paris: Ministère de la Marine.
910.4

New Caledonia

LABILLARDIERE, Jacques Julien Houton de

1800 Vocabulary of the language of the natives of New Caledonia. In his *Voyage in search of La Pérouse during the years 1791, 1792, 1793 and 1794*, vol.II, 409-418. London.
F 307

D'ENTRECASTEAUX, Joseph-Antoine Raymond Bruni

1808 Vocabulaire de la langue des habitans de la Nouvelle-Caledonie. In his *Voyage d'Entrecasteaux, envoyé à la recherche de La Pérouse*, vol.I, 573-584. Paris: Imprimerie Impériale. (Includes some words from Hohoua Island.)
F 461

New Ireland

GAIMARD, M.

1834 Vocabulaire de la langue des habitans du Havre-Carteret (Nouvelle-Irelande). In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie*, vol.II, 143-145. Paris: Ministère de la Marine.
910.4

Rotuman

LESSON, R.P.

1825? Notice sur l'Ile de Rotouma située dans le Grand-Océan Austral. In *Nouvelles annales des voyages de la géographie et de l'histoire*, 159-178. Paris: Libraire de Gide Fils.
N 919.66

1839 Vocabulaire. In his *Voyage autour du monde sur la corvette La Coquille*, vol.II, 439-444. Paris.
F 2787

Vanuatu

GAIMARD, M.

1834 Vocabulaires des idiomes des habitans de Vanikoro. In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie*, vol.II, 165-174. Paris: Ministère de la Marine. (Includes a comparison of the languages of Vanikoro, Tanema, Taneanou.)
910.4

POLYNESIAN LANGUAGES

COOK, James

1773 [Comparison between New Zealand and Otaheite languages]. In John Hawkesworth *An account of the voyages for making discoveries in the southern hemisphere ...*, vol.III, 70-72. London.
RBNq COO

KENDALL, Thomas

1817 Vocabulary of England and New Zealand words, the latter being compared with some of those in the Tongan language. In John Liddiard Nicholas *Narrative of a voyage to New Zealand*, 325-352. London.
F 690

MARSDEN, William

1834 On the Polynesian or East Insular languages. In his *Miscellaneous works*, Part 1. London.
499.1

GAIMARD, M.

1834 Considérations sur les dialectes de la langue polynésienne. In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philolcgie*, vol.II, 263-306. Paris: Ministère de la Marine.
910.4

1834 Vocabulaire comparatif de sept dialectes de la langue océanienne. In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philolcgie*, vol.II, 198-261. Paris: Ministère de la Marine. (Comparison of Malagasy, Malay, Maori, Tongan, Tahitian and Hawaiian.)
910.4

MOERENHOUT, J.A.

1837 Langue; grammaire et littérature des Polynesiens. In his *Voyages aux îles du grand océan ...*, vol.I, 395-415. Paris.
910.8

WILLIAMS, John

1837 Language and dialects of the Polynesians. In his *Narrative of missionary enterprises in the South Sea islands*, Chapter 30. London.
N 266.58 WIL

Hawaiian

ANDERSON, William

1784 Vocabulary of the language of Atooi, one of the Sandwich Islands, January 1778. In James Cook and James King *Voyage to the Pacific ocean for making discoveries in the Northern hemisphere ...*, vol.III, 547-551. London: W. and A. Strahan, Printers.
RBNq 910.4

BERESFORD, William

1789 Vocabulary of 82 words of Sandwich Islands, collected by William Beresford 1786-1787. In George Dixon *Voyage round the world ...*, 268-270. London.
q 910.4

LISIAŃSKII, Iūriĭ Fedorovich

1814 Vocabulary of the language of the Sandwich Islands. In his *Voyage round the world in the years 1803,4,5 and 6 ...*, 326-328. London.
q 910.4

CAMPBELL, Archibald

1816 Vocabulary of the language of the Sandwich Islands. In his *Voyage round the world from 1806 to 1812 ...*, 225-255. Edinburgh. (Includes a vocabulary of 450 words.)
910.4

ARAGO, Jacques Etienne Victor

1822 Vocabulaire: Iles Sandwich. In his *Promenade autour du monde, pendant les années 1817, 1818, 1819 et 1820 ...*, 493-500. Paris.
N 910.4

ELLIS, William

1827 Remarks on the orthography of native words. In his *Narrative of a tour of Hawaii, or Owhyhee: with remarks on the history, traditions, manners, customs and language of the inhabitants of the Sandwich Islands*. London: Fisher, Son and Jackson.
N 266 ELL

ANONYMOUS

1830 *He MAU himeni Hawaii he mea hoolea'i i ke akua mau ia Iehova*. Oahu: Na Misionari i Pai.
499.69

1832 *He NINAUHOIKE no ka mooolelo o ka palapapa* [sic] *hemolele: i ikea ai ka hana a ke akua a me na kanaka, mai ke kumu mai a hiki i ka wa o ka ie mau lunaolelo: he mea e pono ai na makua, a me ne keiki, a me na kula:* 2nd edn. Oahu: Na Misionari i Pai. (A historical catechism of the Holy Scriptures.)
499.69

1832 *O KA buke mua e ao ai i ka heluhelu palapala i ike ai hoi na haumana*. Hawaii: nei i ka Ponemadli. (Language readers.)
SR 499.486 011

1832 *He UI no ka mooolelo kahiko a ke Akua, he mea e ao aku ai i na Kamalii*. Oahu: na Misionari.
SR 499.69

1834 *He MOODELA no na holoholona wawae eha*. Lahainaluna. (Natural history lessons in Hawaiian.)
SR 499.69 HEM

ANDREWS, Lorrin

1836 *Vocabulary of words in the Hawaiian language*. Lahainaluna.
499.69

BIBLE. N.T. Hawaiian

1837 *Ke kauoha Hou a ko kakou Haku e Ola'i a Iesu Kristo: oia ka olelo hemolele no ke ola a na lunaolelo i kakau ai: ua unuhiia mai ka olelo Helene*. Honolulu: Ua paiia na ko Amerika Poe i huiia e hoolaha i ka Baibala, ka na misionari mea pai.
JAF 499.69

CHAMISSO, Adelbert

1837 *Über die Hawaische Sprache*. Leipzig.
SR 499 CHA

ANONYMOUS

1838 *Na HAAWINA kamalii, na mea eao ai na kamalii, ma ke kula Sabati*. Honolulu: Paipalapala a na Misionari. (Catechism for children.)
499.65

BIBLE. O.T. Hawaiian

1838 *KA Palapala hemolele o Iehova ko kakou akua: o ke kauoha kahiko i unuhiia mai ka olelo Hebera*. [Honolulu]: Paiia no ko Amerika Poe Hoolaha Baibala, na na misionari i pai.
499.69

ARAGO, Jacques Etienne Victor

1839 Vocabulaire: Iles Sandwich. In his *Souvenirs d'un aveugle ...*, vol.IV, 402-410. Paris.
F 3547

Maori

PARKINSON, Sydney

1773 A vocabulary of the language of New Zealand. In his *Journal of a voyage to the South Seas in His Majesty's ship the Endeavour.* London.
910.4 f G.M.M.

KING, P.G.

1798 A short vocabulary of the New Zealand language. In David Collins *Account of the English colony in New South Wales ...*, 532-536. London.
F 263

SAVAGE, John

1807 Language, vocabulary, numerals. In his *Some account of New Zealand*, 72-79. London.
NK 387

NICHOLAS, John Liddiard

1817 Three songs in the native language. In his *Narrative of a voyage to New Zealand ...*, vol.I, 69-71. London.
F 690

1817 Vocabulary of the New Zealand language. In his *Narrative of a voyage to New Zealand ...*, vol.II, 323-352. London.
F 690

LEE, Samuel and Thomas KENDALL

1820 *A grammar and vocabulary of the language of New Zealand.* London: Church Missionary Society. Williams, no.2.
499.31

PHILOLOGICUS

1829 The language of New Zealand. In [*Articles on Polynesia*].
919.6

DUMONT D'URVILLE, Jules Sebastian César

1830 Langage et numération des Nouveaux-Zélandais. In his *Voyage de la corvette l'Astrolabe ...*, vol.II, 563-570. Paris: J. Tastu.
F 1341

YATES, William

1830 *Ko te Pukapuka Tuatahi o Mohi e huaina ana ko Kenehi.* Sydney: R. Mansfield, Printer. (Scripture portions, prayers, catechism and hymns.) Williams no.6.
F 1378

'SOLICITUS'

1831 Origin, language and religion of the New Zealanders. *Sydney Gazette* 8 January:3. (Letter to the Editor.)
mfm Nx35

ANONYMOUS

1833 [Lesson sheets in Maori]. Sydney: Jones. (Four sheets containing Maori alphabet, scripture texts, prayers for children and a notice.) Williams no.12.
NK 993

1833 *Ko te Pukapuka Inoinga, me nga Karakia Hakarameta, me era Ritenga hoki o te Hahi o Ingarani*. Hirini: Kua oti te ta e te Tipene raua ko te toki. (Morning and evening service, sacraments, catechisms, and hymns.) Williams no.9.
F 1672

BIBLE. Selections. Maori

1833 *Ko te tahi wahi o te Kawenata Hou o Ihu Karaiti te Ariki, to tatou kai wakaora: me nga upoko e waru o te Pukapuka o Kenehi*. Hirini: Kua oti te ta e Te Tipene raua ko Te Toki. Williams no.8.
NK 1438

GAIMARD, M.

1834 Vocabulaire Français-Mawi (Nouvelle-Zélande): vocabulaire Mawi-Français. In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie*, vol.II, 14-55. Paris: Ministère de la Marine.
910.4

BIBLE. N.T. Epistles. Maori

1835 *Ko nga Pukapuka o Paora te Apotoro ki te Hunga o Epeha o Piripai*. Paihia: He mea ta i te Perehi o nga Mihanere o te Hahi o Ingarani. Williams no.15.
499.31

BIBLE. N.T. Maori

1837 *Ko te Kawenata Hou o to tatou Ariki te kai wakaora a Ihu Karaiti. He mea wakamaori i te reo kariki*. Paihia: He mea ta i te Perehi o te Hahi o Ingarani. (Translated by William Williams.) Williams no.20.
JAF 499.4 B582

ANONYMOUS

1838 *KO nga Ritenga mo te Hunga e Huihui ana ki nga Karahi*. New Zealand: He mea ta i te Perehi o nga Mihanere Weteriana. Williams no.27a.
NK 2531

1838 *He PUKAPUKA Wakaako*. Mangungu [N.Z.]: He mea ta i te Perehi o nga Mihanere Weteriana. Williams no.28.
NK 998

BROUGHTON, William Grant

1839 *Ko te Pukapuka Kauwau o te Pihopa, ki te Hunga Wakapono o Nu Tirani, e huihui ana ki te Hahi o Ingarani, te Wakaminenga ki Paihia, mo te Karakia o te Wakapanga ringaringa Hanueri 5*. Paihia: He mea ta i te Perehi o nga Mihanere o te Tahi o Ingarani. Williams no.33.
499.31P

BUMBY, John Hewgill

1839 *Ko te Pukapuka Aroha o nga Kaumatua o te Hahi Weteriana o Ingarani, Matou ko nga Mihanere, ki o matou tangata i Nu Tirani*. Mangungu [N.Z.]: He mea ta i te Perehi o nga Mihanere Weteriana o Ingarani. (A pastoral letter signed by Bumby from the Wesleyan elders and missionaries to their converts.) Williams no.36a.
NK 2535

Marquesan

LANGSDORFF, *Baron* Georg Heinrich von

1812 Spracheproben von der Insel Nukahiwa. In his *Bemerkungen auf einer Reise um die Welt in den Jahren 1803 bis 1807*, vol.1, 153-159. Frankfurt.
910.4

LISIAŃSKIĬ, Iūriĭ Fedorovich

1814 Vocabulary of the language of Noocahiva, Marquesas Islands. In his *A voyage round the world in the years 1803, 4, 5 and 6 ...*, 323-325. London.
q 910.4

SHILLIBEER, John

1817 List of words, numerals and phrases of the inhabitants of Port Anna Maria, Nooaheevah, Marquesas Islands, 1814. In his *Narrative of the Briton's voyage to Pitcairn's island*, 57-61. London.
919.7

Rarotongan

BIBLE. N.T. Rarotongan

1836 *Te korero-motu ou a to tatou atu e te ora a Jesu Mesia: kiritiia i te reo Rarotonga*. Lonedona: British and Foreign Bible Society.
JAF 499.623

ANONYMOUS

1837 *O le UPU ia losefu, na faamatalaina o le upu Samoa*. Rarotonga: Mission Press.
SR JAF 499.623

BIBLE. O.T. Pentateuch. Rarotongan

1838 *Te au buka a Mose: kiritiia ei tuatua Rarotonga*. Rarotonga: Misionari Societi i Lonedona.
JAF 499.623

Samoan

ANONYMOUS

1836 *E mou IMENE, o lea foi le tala faalelei i le Atua*. Huahine: [Mission Press].
SR 499.613

BIBLE. N.T. Matthew. Samoan

1836 *O le evagelia a Mataio i le Mesia ia Iesu lo tatou alii, ua litiu i le upu Samoa*. Huahine: Mission Press.
SR 499.613

BIBLE. O.T. Genesis. Samoan

1837 *Le upu a le mou tagata anamua, ma Atamo, ma Aperaamo e ma tagata tele uma, ua na liu i le upu Samoa*. Huahine: Mission Press.
SR 499.613

Tahitian

BANKS, *Sir* Joseph

1772 Vocabulaire abrége de la langue de l'île Otahiti. In *Supplément au voyage de M. de Bougainville; ou Journal d'un voyage autour du monde,*

fait par MM Banks & Solander ... traduit de l'anglois par M de Fréville, 241-250. Paris: chez Saillant & Nyon.
RB 910.4 BOU

BOUGAINVILLE, Louis Antoine de

1772 Vocabulary of the language of Taiti Island. In his *A voyage round the world ... translated from the French by John Reinhold Forster*, 470-476. London: J. Nourse and T. Davies, Printers.
910.4 BOU

1773 Vocabulaire de l'Isle Taiti. In his *Voyage autour du monde par le frégate du roi La Boudeux, et la flûte l'Etoile en 1766-69*. 2nd edn. Neuchatel.
910.4

COOK, James

1773 Language [Tahitian]. In John Hawkesworth *An account of the voyages ... performed by Commodore Byron, Captain Wallis, Captain Carteret and Captain Cook ...*, vol.II, 228-230. London: W. Strahan and T. Cadell, Printers.
RBNq COO

PARKINSON, Sydney

1773 A vocabulary of the language of Otaheiti. In his *Journal of a Voyage to the South seas ...*, 51-66. London.
RBNq 910.4

ANONYMOUS

1774 VOCABULAIRE de l'isle Taiti. In M. Richer *Histoire moderne ...*, vol.25, 437-455. Paris: Saillant and Nyon; Veuve Desaint.
RB 909 MAR

COOK, James

1777 A vocabulary of the language of the Society Isles. In his *A voyage towards the South Pole and round the world*, vol.II, 317-363. London.
RBN 910.4

ANONYMOUS

1814 *PARAU no Iesu Christ te Temaidi no te Atua: e no te mou pipi nona*. Sydney: G. Howe, Printer.
F 594

BIBLE. N.T. Luke. Tahitian

1818 *Te Evanelia na Luka, iritihia ei parau Tahiti*. Moorea.
SR 499.6211 TEE

BIBLE. N.T. Matthew. Tahitian

1820 *Te Evanelia a Mataio, no Iesu Christ to tatou fatu; iritihia ei parau Tahiti*. Tahiti: printed at the Windward Press.
SR 499.6211 TEE

BIBLE. Selections. Tahitian

1822-1828 *Otaheitan*. Tahiti: Mission Press, London Missionary Society.
499.6211

BIBLE, N.T. Acts. Tahitian

1822 *Te ohipa a te mau aposetolo, i papai hia e Luka, iriti hia ei parau Tahiti*. Tahiti: [Mission Press, London Missionary Society]. (Translated by Henry Nott and Pomare II.)
JAF 499.6211

DAVIES, John

1822 *Aritemeti oia te haapaoraaotetai o e te faa au raa o te numera.* [Tahiti: Windward Mission Press, London Missionary Society].
SR 499.6211

ANONYMOUS

1822 *E PARAU faaitoito i te tamarii rii, e te taata paari atoa hoi.* Tahiti: [Mission Press, London Missionary Society].
SR 499.6211

BIBLE. N.T. Acts. Tahitian

1823 *Te ohipa a te mau aposetolo na Iesu Christ te fatu; iritihia ei parau Tahiti.* Tahaa: Leeward Mission Press, [London Missionary Society].
JAF 499.6211

BIBLE. N.T. Epistles. Tahitian

1824 *Te mau episetole a te aposetolo ra a Paulo, ta'na i papai adu i to Galatia, Ephesia, Philipi, Colosa, Tesalonia, ia Timoteo hoi, Tito e Philemona.* Tahiti: [Windward Mission Press, London Missionary Society]. (Translated by John Davies.)
JAF 499.6211

BIBLE. O.T. Selections. Tahitian

1824 *Te buka a te peropheta a Daniela, e te buka, hoi a Ruta e a Eseta, iritihia ei parau Tahiti.* Tahaa: Leeward Mission Press, [London Missionary Society]. (Translated by John Williams.)
JAF 499.6211

BIBLE. N.T. Epistles. Tahitian

1825 *Te mau episetole a te aposetolo ra a Paula, tana i papai adu i to Galatia, Ephesia, Philipi, Colosa, Tesalonia, ia Timoti hoi, ia Tito, e ia Philemona, iritihia ei parau Tahiti.* Tahaa: Leeward Mission Press, [London Missionary Press].
JAF 499.6211

LONDON Missionary Society

1825 *E parau faaitoito i te Societi Misionari i Tahiti e i Moorea atoa hoi.* [Tahiti: Mission Press]. (Report of the Windward Division of the Tahitian Mission.)
SR 499.6211

MAI, Chef de Bora Bora

1825 [Letter]. MS, Manuscripts Section, National Library of Australia. (Written in Tahitian to the Directors of the London Missionary Society, London.)

BIBLE. N.T. Selections. Tahitian

1826 *Te episetole i te mau Hebera ra, e ta Iacobo, te mau episetole a Ioane, e te Apocalupo hoi, iritihia ei parau Tahiti.* Tahaa: Leeward Mission Press, [London Missionary Society].
SR 499.6211

ANONYMOUS

1826 *Te REPOSITORI, oia hoi to pue raa parau maitatoi no.1.* Tahiti: Mission Press, [London Missionary Society]. (Mainly a Tahitian translation of a letter, written in 1825 to the missionaries in the Society Islands from Daniel Tyerman and George Bennett.)
SR 499.6211

ANONYMOUS

1827 *E BUE raa himene, oia hoi te parau haamaitai te Atua.* Tahiti: Mission Press, [London Missionary Society]. (Index to Tahitian hymn-book *E faaite, ia itea i te araara o te mau himene atoa nei.*)
SR 499.6211

BIBLE. N.T. Mark. Tahitian

1827 *Te evanelia a Mareko, iritihia ei parau Tahiti.* Tahiti: [Mission Press, London Missionary Society]. (Translated by John Davies.)
JAF 499.6211

TAERO, Habitant de Maupiti

1827 Lettre en tahitien signée au Rev. George Burder, Mission House, Austin Friars, London. MS, Manuscripts Section, National Library of Australia.
MS 2639

WESTMINISTER Assembly of Divines

1827 *E ui tumu no te mau parau a te Atua; e tia ia faaroo paaatoahia e te taata toa nei.* Tahiti: [Mission Press, London Missionary Society]. (Catechism.)
SR 499.6211

BIBLE. N.T. Epistles. Tahitian

1828 *Te episetole a te aposetolo ra a Paulo, i to Roma, e na episetole epiti hoi i to Corinetia, na episetole a Petero e te episetole a Iuda.* Tahiti: [Mission Press, London Missionary Society]. (Translated by John Davies.)
JAF 499.6211

ANONYMOUS

1830 BORA Bora [Letter] to Na Temau Misionari Societi, Paratane, London. (Written in Tahitian by Mai?)
MS 2641

BIBLE, O.T. Tahitian

1832 *Te buka salamo te rahi i Papaihia e Davida.* Tahiti.
SR 499.6211 BIB

BIBLE. O.T. Isaiah. Tahitian

1833 *Te buka a te peropheta ra a Isaiah: iritihia ei parau Tahiti* ... Huahine: Mission Press.
499.62

BIBLE. O.T. Minor prophets. Tahitian

1834 *Te mau buka a te tahi pae peropheta ra ... iritihia ei parau Tahiti.* Tahiti: Windward Mission Press.
499.62

MOERENHOUT, Jacques Antoine

1837 Langue, littérature. In his *Voyages aux îles du grand Océan* ... 395-418. Paris.
919.6

BIBLE. N.T. Tahitian

1838 *Te faufaa api, a to tatou fatu e te ora a Iesu Mesia Ra; iritihia ei parau Tahiti.* Lonedona, Neia e Richard Watts.
499.62

NOTT, Henry, et al, trans.

1838 *Te Bibilia moa ra, oia te faufaa tahito e te faufaa api ra: iritihia ei parau Tahiti.* Lonedona: Neia e Richard Watts, na te Taieti Bibilia i Beritani e te mau fenua toa i faatupu hi i Lonedona ra i te Matahiti 1804.
NK 999

Tongan

LABILLARDIERE, Jacques Julien Houton de

1800 Vocabulaire de la langue des îles des Amis. In his *Relation du voyage à la recherche de La Pérouse ...* Tome second, 47-57. Paris: H.J. Jansen.
F 307

LE MAIRE, Jacques

1806 Vocabulary of the language of the Cocos Islands. In James Burney *A chronological history of the discoveries in the South Sea or Pacific Ocean*, vol.II, 440-442. London: Luke Hansard, Printer.
F 372

D'ENTRECASTEAUX, Joseph-Antoine Raymond Bruni

1808 Vocabulaire de la langue des habitans des îles des Amis. In his *Voyage d'Entrecasteaux, envoyé à la recherche de La Pérouse ...*, 557-572. Paris: Imprimerie Impériale.
F 461

MARINER, William

1818 *An account of the natives of the Tonga islands; with an original grammar and vocabulary of their language by John Martin.* 2nd edn. London.
919.612

TURNER, Nathaniel

1828 *First lessons in the language of Tongataboo, one of the Friendly Islands; to which are added, a prayer and several hymns.* Sydney: R. Howe, Government Printer.
F 1221

METHODIST Mission, Nukualofa, Tonga Islands

1832 *Koe gaahi lesoni eni, mei he Tohi Tabu.* [Nukualofa, Tongataboo]. (Bible lessons.)
JAF 499.612

1833 *Koe gaahi lesoni mei he Tohi-tabu: no.9.* Tongataboo: Printed at the Mission Press.
Np 499.61

BIBLE. O.T. Isaiah. Selections. Tongan

1834 *Koe buka ae balofite ko Aisea.* [Tongataboo: Wesleyan Mission].
JAF 499.612

BIBLE. O.T. Psalms. Selections. Tongan

1834 *Koe Same; no.15.* Tonga: Mission Press.
Np 499.61

BIBLE. O.T. Samuel. Selections. Tongan

1834 *Koe uluaki buka a Samiuela.* [Tonga: Wesleyan Mission].
JAF 499.612

GAIMARD, M.

1834 Vocabulaire Français-Tonga; vocabulaire Tonga-Français. In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie*, vol.II, 56-136. Paris: Ministère de la Marine.
910.4

METHODIST CHURCH IN THE TONGA ISLANDS

1834 *Koe fehui moe tala, mei he uluaki folafola ae otua.* [Tonga].
JAF 499.612

BIBLE. N.T. Matthew. Selections. Tongan

1835 *Matiu.* [Feletoa, Vavau: Wesleyan Mission]. (Translated by John Thomas.)
JAF 499.612

ANONYMOUS

1837 *KOE fehui moe tala, mei he Uluaki folofola ae otua.* Vavau: Mission Press. (Catechism.)
JAF 499.612

BIBLE. N.T. Acts. Selections. Tongan

1837 *Koe gaahi gaue ae kau Abosila.* [Vavau: Wesleyan Mission].
JAF 499.612

BIBLE. N.T. Epistles. Tongan

1837 [The Epistles of John and Jude, translated into Tongan]. [Vavau: Wesleyan Mission].
JAF 499.612

BIBLE. N.T. Epistles. Selections. Tongan

1838 [The Epistles of Paul to the Thessalonians, Timothy, Titus and Philemon, translated into Tongan]. [Vavau: Wesleyan Mission].
JAF 499.612

BIBLE. N.T. John. Tongan

1838 *Koe kosibeli nae tohi e Jone.* [Vavau: Wesleyan Mission].
JAF 499.612

BIBLE. O.T. Genesis. Tongan

1838 *Koe uluaki euka a Mosese oku ui ia ko Jenesi.* [Vavau: Wesleyan Mission].
JAF 499.612

Other Polynesian languages

LE MAIRE, Jacques

1806 Vocabulary of the language of the Solomon Islands, i.e. the Hoorn Islands 1616. In James Burney *A chronological history of the discoveries in the South Sea or Pacific Ocean*, vol.II, 439-440. London: Luke Hansard, Printer.
F 372

GAIMARD, M.

1834 Vocabulaire de la langue des habitans de Tikopia. In J.S.C. Dumont D'Urville *Voyage de découvertes de l'Astrolabe ... Philologie*, vol.II, 161-164. Paris.
910.4